

General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.



MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROGRESS REPORT: THE ANALYSIS OF THE PILOT'S
COGNITIVE AND DECISION PROCESSES

NGR 22-009-733

March 1, 1975 - August 31, 1975

(NASA-CR-145739) THE ANALYSIS OF THE
PILOT'S COGNITIVE AND DECISION PROCESSES
Progress Report, 1 Mar. - 31 Aug. 1975
(Massachusetts Inst. of Tech.) 85 p HC
\$5.00

N76-11722

Unclas

CSCL 05E G3/53 01930

PRINCIPAL INVESTIGATOR

R.E. CURRY



MAN-VEHICLE LABORATORY
DEPARTMENT OF AERONAUTICS AND ASTRONAUTICS
CENTER FOR SPACE RESEARCH
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139

PROGRESS REPORT

THE ANALYSIS OF THE PILOT'S COGNITIVE AND DECISION PROCESSES

MAN-MACHINE INTEGRATION BRANCH

NASA AMES RESEARCH CENTER

NGR 22-009-733

March 1, 1975 - August 31, 1975

Professor R.E. Curry
Man-Vehicle Laboratory
Department of Aeronautics
and Astronautics
Massachusetts Institute of
Technology

TABLE OF CONTENTS

Progress Report 1

Appendices

Appendix A

Detection of System Failures in Multi-Axis Tasks

Appendix B

Changes in Pilot Workload During an Instrument Landing

Appendix C

Failure Detection by Pilots During Automatic Landing:
Models and Experiments

Appendix D

Experiments in Pilot Decision-Making During Simulated
Low Visibility Approaches

Appendix E

A Multinomial Maximum Likelihood Program (MUNOML)

Appendix F

A Random Search Algorithm for Laboratory Computers

One of the major accomplishments during this reporting period was the completion of the year's second doctoral thesis. Dr. Arye R. Ephrath completed his degree requirements in June of 1975 and his thesis, abstracted here, has been made into a NASA Contractor's Report #CR-137759.

PILOT PERFORMANCE IN ZERO-VISIBILITY PRECISION
APPROACH

Arye R. Ephrath

PH.D. Thesis, Department of Aeronautics and Astro-
nautics, Massachusetts Institute of Technology, June
1975

ABSTRACT

To gain a better understanding of the pilot's short term decisions regarding performance assessment and failure monitoring, the performance of fifteen air line pilots who flew simulated zero-visibility landing approaches is reported. The approaches were flown with different degrees of automation and at different workload levels, as induced by simulated wind disturbances and measured by a sensitive, non-loading subsidiary task.

The results indicate that the pilot's mode of participation in the control task has a strong effect on his workload, the induced workload being lowest when the pilot acts as a monitor during a coupled approach and highest when the pilot is an active element in the control loop. In addition, a very marked increase in workload at altitudes below approximately 500 feet AGL is documented at all participation modes; this increase is inversely related to distance-to-go.

The effects of workload and participation mode on failure detection are separated. The participation mode is shown to have a dominant effect on failure

2.


detection performance, with a failure in a monitored (coupled) axis being detected faster than a comparable failure in a manually controlled axis.

Touchdown performance is also documented and the findings of previous investigators are supported, namely that the conventional instrument panel and its associated displays are quite inadequate for zero-visibility operations in the final phases of the landing approach.

A paper summarizing some of the work presented in this thesis was delivered by Dr. Ephrath at the Eleventh Annual Conference on Manual Control. The paper, entitled "Detection of System Failures in Multi-Axes Tasks", is contained in Appendix A.

A research note in preparation by Dr. Curry and Dr. Ephrath is contained in Appendix B. Its title is "Changes in Pilot Workload During an Instrument Landing"

A paper on modelling pilot performance in failure detection tasks was presented at the Eleventh Annual Conference on Manual Control by Drs. Curry and Gai. The paper has since been revised and submitted for publication. The abstract of the paper which was presented is given below and the text of the paper is given in Appendix C.



FAILURE DETECTION BY PILOTS DURING AUTOMATIC LANDING:
MODELS AND EXPERIMENTS

Eli G. Gai and R.E. Curry

ABSTRACT

A model of the pilot as a monitor of instrument failures during automatic landing is proposed. The failure detection model of the pilot consists of two stages: a linear estimator (Kalman Filter) and a decision mechanism which is based on sequential analysis. The filter equations are derived from a simplified version of the linearized dynamics of the aircraft and the control loop. The perceptual observation noise is modelled to include the effects of the partition of attention among the several instruments. The final result is a simple model consisting of a high pass filter to produce the observation residuals, and a decision function which is a pure integration of the residuals minus a bias term.

The dynamics of a Boeing 707 were used to simulate the fully coupled final approach in a fixed base simulator which also included failures in the airspeed, glideslope, and localizer indicators. Subjects monitored the approaches and detected the failures; their performance was compared with the predictions of the model with good agreement between the experimental data and the model.

During this reporting period, the data collected during the principal investigator's year at Ames Research Center was analyzed. A discussion of the experiments performed, the data analysis, and the results of these experiments are contained in a paper presented at the Eleventh Annual Conference on Manual Control. The paper is abstracted here and presented in full in Appendix D.

4

EXPERIMENTS IN PILOT DECISION-MAKING DURING SIMULATED LOW VISIBILITY APPROACHES

R.E. Curry, J.K. Lauber, & G.E. Billings

ABSTRACT

Despite a vast accumulation of operational experience with the conduct of low visibility instrument approaches, little is understood about the decision making behavior of pilots who fly these approaches. Likewise, there is little information regarding the man, system and task related factors which influence this decision-making behavior. Such information is essential for the rational design of new systems, or for the redesign of existing systems in order to correct known deficiencies.

A major problem which has inhibited the study of pilot decision-making behavior in the laboratory has been the unavailability of tasks which incorporate the essential cognitive features of the real task, and which include those motivational or stress-inducing features known to influence decision-making performance. This paper describes a task which was designed to simulate the cognitive features of low visibility instrument approaches, and to produce controlled amounts of subjective stress in pilots serving as subjects in experiments using the task.

Pilot behavior during low visibility instrument approaches can be analyzed into at least two major categories: one is the continuous closed-loop manual tracking behavior necessary to control the aircraft, and the other is the cognitive, decision-making behavior required to make the decision to continue the approach and landing or to execute a missed approach. It is the second category of behavior with which we are concerned here.

The difficulty of a decision making task is, in part, determined by the uncertainty of the data used to make the decision. For example, the decision to "go-around" is a relatively easy one if, at the missed approach point, there is nothing to be seen outside the aircraft, or if the approach lights and runway have been clearly visible for the last two miles of the approach. It is when the approach lights or runway are barely visible, and then only intermittently, that the decision becomes more difficult.

5

A second class of variables which are known to influence the outcome of decision-making tasks is best illustrated by the various kinds of psychological stressors acting upon the pilot. Of particular interest here are the pressures perceived by the pilot to complete the approach, to make an on-time arrival, to save fuel, and even to save "face".

We have assumed that it is necessary to use a simulation task which incorporates both kinds of variables, informational and psychological, to successfully study pilot decision-making behavior in the laboratory. The task discussed below was designed to meet those requirements. This paper describes our preliminary experiments in the measurement of decisions and the inducement of stress in simulated low visibility approaches.

PUBLICATIONS

Two major publications resulting from this research appeared during the reporting period. Both appeared in Behavior Research Methods and Instrumentation and they are abstracted below. Full copies of the papers appear in the Appendices E and F. A more detailed description of MUNOML containing program listings and subroutines is available from the principal investigator.

6

A MULTINOMIAL MAXIMUM LIKELIHOOD PROGRAM (MUNOML)

R.E. Curry

Behavior Research Methods and Instrumentation,
Volume 7(3), pages 305-307, 1975.

ABSTRACT

In our research on modelling sensory and decision phenomena, we were soon confronted with the task of evaluating both old and new models using both old and new data. Rather than design an ad-hoc estimation program for each new model, as is typically done, we developed an "executive" program that provides a general method for estimating parameters and simultaneously provides flexibility for accommodating new models with a minimum amount of programming. Our experience with canned computer programs has been equivocal, so we decided to provide only the general framework and let the user accomplish the objective of estimating the parameters for his particular model by writing a new subroutine within the constraints of the executive program.

The most common class of distributions for which parameters must be extracted are multinomial distributions resulting from a stimulus-response classification, e.g. binary responses (YES-NO or two alternative forced choice methods), the method of successive categories (rating scales), or transition probabilities in a Markov chain. Although a number of methods exist for estimating such parameters, we have chosen the maximum likelihood method and have implemented the scoring of Rao to adjust the parameters from one iteration to the next. We have chosen the maximum likelihood (ML) method because (1) it is a member of the class of consistent asymptotically normal estimators (CAN) under suitable conditions; (2) it will easily handle situations in

which all the responses fall into one category;
(3) there are many situations in which the maximum likelihood estimator can be shown to yield unique estimates for parameters (other estimation techniques may have this property as well); and (4) it is the only one exhibiting first order efficiency.

Other authors have successfully used the maximum likelihood method with Rao's scoring method in signal detection type models. This program differs primarily in its ability to estimate parameters for a wide variety of models.


A RANDOM SEARCH ALGORITHM FOR LABORATORY COMPUTERS

Renwick E. Curry

Behavior Research Methods and Instrumentation,
Volume 7(4), pages 369-376, 1975.

ABSTRACT

The small laboratory computer is ideal for experimental control and data acquisition. Postexperimental data processing is many times performed on large computers because of the availability of sophisticated programs, but costs and data compatibility are negative factors. Parameter optimization, which subsumes curve fitting, model fitting, parameter estimation, least squares, etc., can be accomplished on the small computer and offers ease of programming, data compatibility, and low cost as attractive features. A previously proposed random search algorithm ("random creep") was found to be very slow in convergence. We present a new method (the "random leap" algorithm) which starts in a global search mode and automatically adjusts step size to speed convergence. A FORTRAN executive program for the random leap algorithm is presented which calls a user supplied function subroutine. An example of a function subroutine is given which calculates maximum likelihood estimates of receiver operating characteristic parameters from binary response data. Other applications in parameter estimation, generalized least squares, and matrix inversion are discussed.



APPENDIX A

PRECEDING PAGE BLANK NOT FILMED

DETECTION OF SYSTEM FAILURES IN MULTI-AXES TASKS

by Arye R. Ephrath

Man-Vehicle Laboratory

Massachusetts Institute of Technology

SUMMARY

The investigation has examined the effects of the pilot's participation mode in the control task on his workload level and failure-detection performance during a low-visibility landing approach. We found that the participation mode had a strong effect on the pilot's workload, the induced workload being lowest when the pilot acted as a monitoring element during a coupled approach and highest when the pilot was an active element in the control loop.

The effects of workload and participation mode on failure detection were separated. The participation mode was shown to have a dominant effect on the failure detection performance, with a failure in a monitored (coupled) axis being detected significantly faster than a comparable failure in a manually-controlled axis.

INTRODUCTION

In the last decade, a great deal of thought has been given to Category III landings and their implications. One area of intensive investigation centers around the role of the crew during the approach, and current thought is polarized around two extremes:

- a. The crew is in the control loop and flies the aircraft in accordance with instrument-generated steering signals.
- b. Steering signals are coupled directly into the autopilot, with the crew monitoring the system.

It is axiomatic that a pilot should be capable of detecting and identifying failures in the automatic landing system accurately, reliably and with minimal time delay. To this end, extensive studies have been conducted in which the pilot was treated as a controlling element in a one-dimensional task; his decision processes (Schrenk, 1969) and his adaptive behavior following a sudden change in the controlled plant dynamics were investigated (Young *et al*, 1964; Phatak and Bekey, 1969). Other studies investigated the failure-detection performance treating the operator as a pure monitor (Gai and Curry, 1975). In reality, however, the pilot is faced with multi-axes, not single-axis, tasks; although models for interference among multiple control tasks have been derived (Levison, 1970), the interrelationships among simultaneous control and monitoring tasks are not yet well understood (Levison, 1971).

Young *et al* (*op. cit.*) found that in single-axis tracking tasks the human operator's performance as a failure detector was better when he was in the control loop; simulated Category III landing studies, on the other hand, have shown that the pilot's failure detection performance deteriorated when he was faced with manual control task, compared to the monitoring mode (Vreuls *et al*, 1968). When faced with split-axis tasks, pilots' monitoring and decision making were impaired (Monroe *et al*, 1968) and they sometimes completely overlooked the occurrence of a failure, presumably because of the increased workload associated with split-axis tasks (Gainer *et al*, 1967).

It has been recognized that when the the role of the human changes from monitoring to that of an active controller corresponding changes take place in his workload level (Ekstrom, 1962; Wewerinke). However, in pilot-performance studies to date these effects were completely confounded. It is the primary purpose of this investigation to separate these effects and to document pilot performance during a Category III landing as a function of the particular control mode at different workload levels. We wished to isolate and identify the effects on performance due to the variations in the control mode alone - and hence, variations in the operator's mode of behavior - apart from the effects on performance due to the variations in the workload level.

METHOD

As stated, the purpose of this research was the study of the pilot's short-term decisions regarding performance assessment and failure monitoring. We wished to investigate the relationship between the pilot's ability to detect failures, his degree of participation in the control task and his over-all workload level. Also, we wished our findings to be applicable to the general population of pilots who fly low-visibility approaches in commercial jet transport aircraft. To this end, this research consisted of an experimental investigation which was carried out in a static ground simulator and which utilized fifteen airline pilots as subjects.

The simulation capability included the ADAGE AGT/30 digital graphics computer and a fixed-base cockpit simulator. A mathematical model has been developed of a large transport aircraft in the landing approach flight envelope; the actual flight data of a DC-8 were used in the equations of motion, and the various parameters were later refined following a series of flight tests by a senior airline captain with considerable Boeing 707/123 experience. Non-linear phenomena such as ground-effect and stalls have also been included.

An integrated-cue flight director system has been designed for this simulator, providing the capability to land the simulated aircraft manually in zero-zero conditions in a relatively satisfactory manner. Also, a two-axis autopilot has been incorporated into the simulation which was capable of intercepting and tracking the Instrument Landing System (ILS), in either axis or in both axes, to touchdown. We also had the capability to add wind disturbances to the simulation to induce different workload levels. The wind gusts were modelled as filtered white noise with a cutoff frequency of $\pi/6$ rad/sec.

The mathematical model was programmed into the AGT/30 computer which was linked via multiplexer channels and sense lines to the cockpit simulator. The cockpit was a mock-up of the captain's crew station in a Boeing transport aircraft (Fig. 1). The windows were frosted to eliminate external visual ref-

ORIGINAL PAGE IS
OF POOR QUALITY

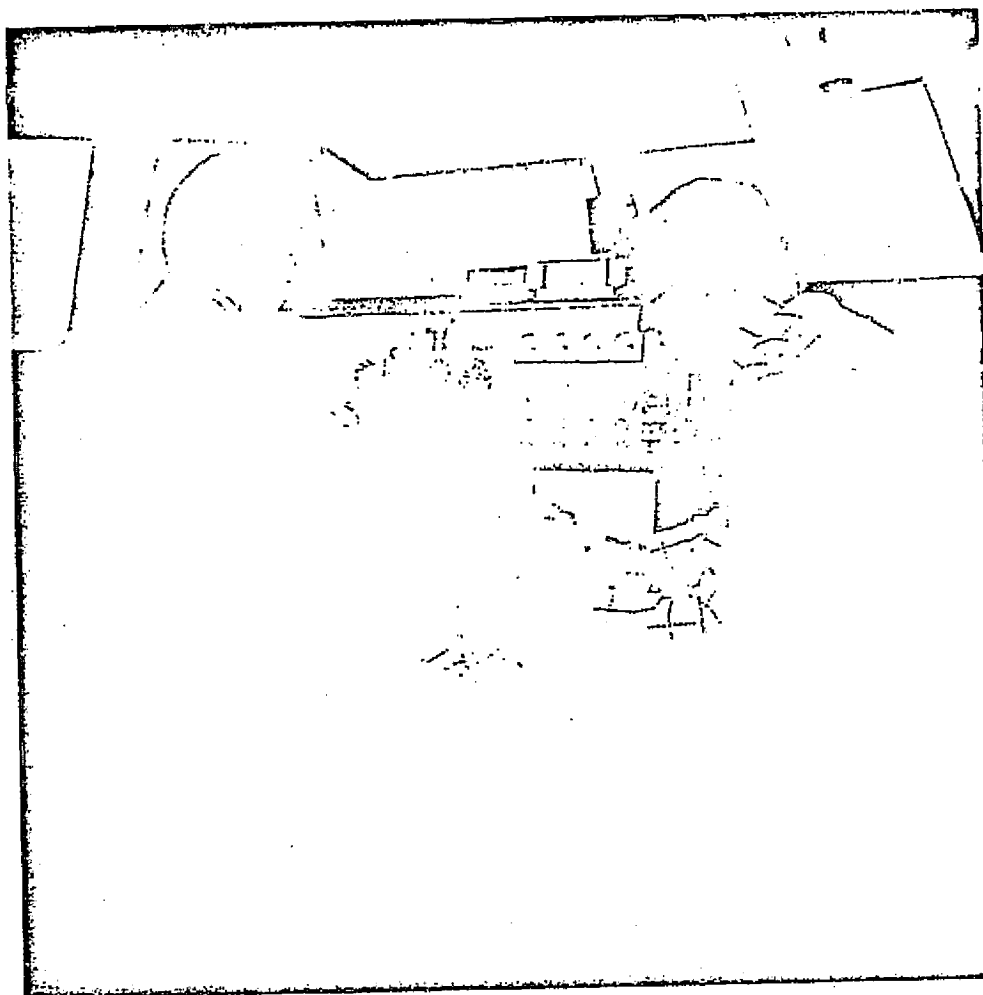


Figure 1. Cockpit Simulator

erence.

The controls included an operational, spring-centered control column with a control wheel and rudder pedals, as well as four throttles, flaps, speed-brake and landing-gear levers and flight-director and autopilot controls.

Apart from engine instruments and marker-beacon lights the simulator was equipped with three CRT screens, mounted one each on the main instrument panel at the captain's and the first officer's stations and one in place of the weather radar screen. The screens were driven simultaneously by the ADAGE computer and presented the six standard flight instruments (Fig. 2): Airspeed, attitude-flight director indicator, altimeter, instantaneous vertical speed indicator, horizontal situation (HSI) and radio-magnetic (RMI) indicators, as well as a DME digital readout and glideslope deviation and course deviation needles. The CRT screens were driven by the computer at a rate of 24 frames per second which was sufficient to produce flicker-free images. The information was updated at a rate of 5/second.

To measure the pilot's workload, a "warning light"-type subsidiary task was selected for the research. It consisted of two small red lights mounted above each other outside the subject's peripheral vision field, and a rocker thumb switch mounted on the left horn of the control yoke.

The lights provided the stimuli. During the run the upper or lower light, with equal probability, was lit at a random time for a maximum of two seconds. A correct response by the subject consisted of turning the light off by a proper motion of the rocker thumb switch. The program recorded the number of times that the subject responded correctly to the warning light ("hits") and his response time (latency) for each response. Incorrect responses by the pilot, that is, not responding to an illuminated light or activating the switch the wrong way, were also counted and labeled as "misses".

A workload index was computed from these data as follows:

- a. As each stimulus was presented for a maximum of 2 seconds, the total response-time ratio RTR for both "hits" and "misses" was computed by

$$RTR = \frac{\text{cumulative latency } (\sum T_i)}{\text{Total number of stimuli} \times 2 \text{ sec}} \quad (1)$$

ORIGINAL PAGE IS
OF POOR QUALITY

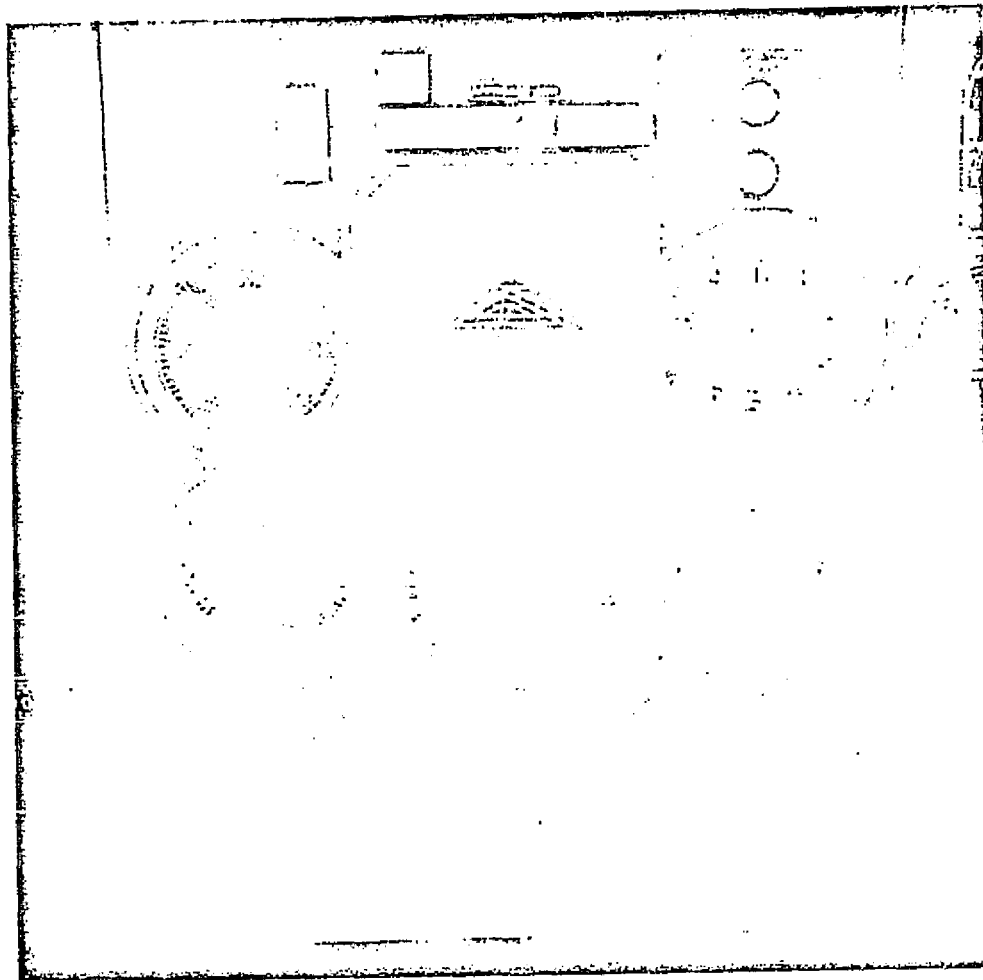


Figure 2. Instrument Panel

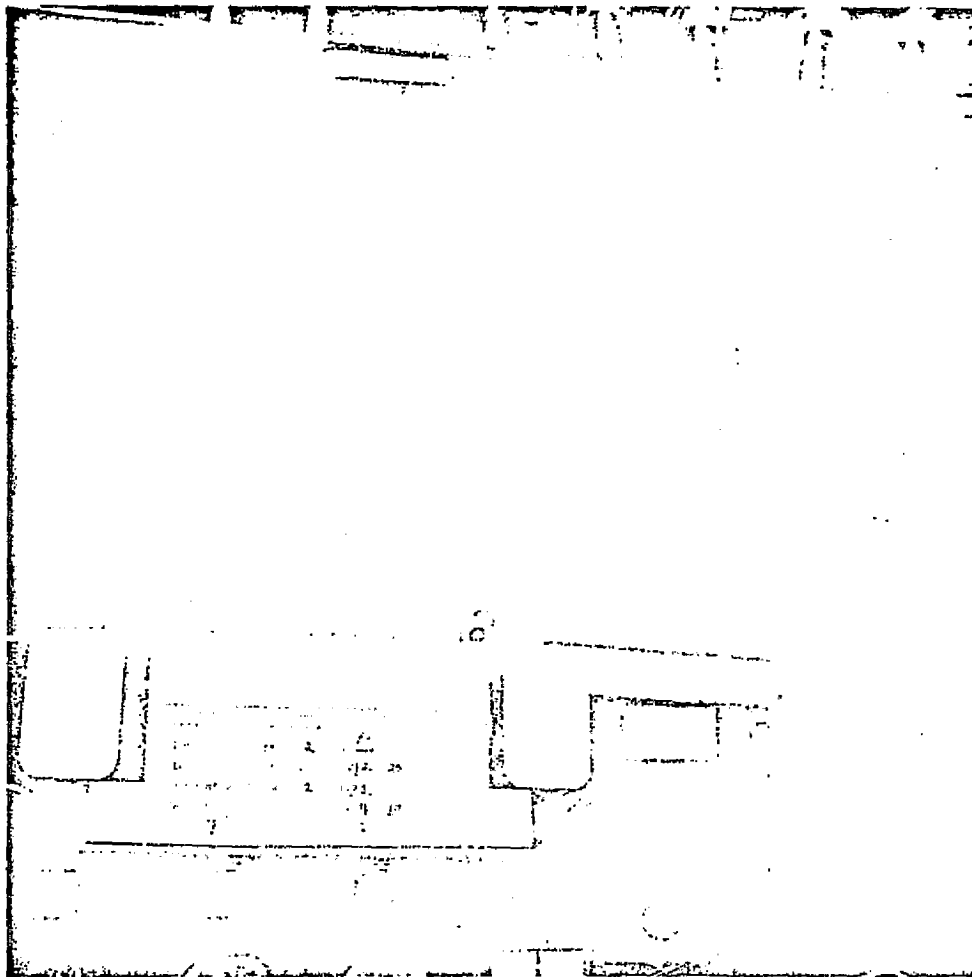


Figure 3. Side-Task Stimulus Lights

ORIGINAL PAGE IS
OF POOR QUALITY

b. A miss-rate MR was computed by

$$MR = \frac{\text{Number of stimuli missed}}{\text{Total number of stimuli}} \quad (2)$$

c. A workload index WLX was then extracted using the best least-squares fit weighing coefficients

$$WLX = \frac{0.78 \text{ RTR} + 0.626 \text{ MR}}{0.78 + 0.626} \times 100 \text{ percent} \quad (3)$$

This measure of workload has been shown (Spyker *et al.*, 1971) to be correlated with physiological predictors of workload with a correlation coefficient $\rho = 0.646$, significant at the $P < 0.005$ level.

d. Finally, we wished to eliminate differences between subjects which may have been caused by different subjects assigning different relative priorities to the primary tracking task and the subsidiary task. To this end the workload index of each subject was normalized, that is, a workload index of zero was assigned to the approach which resulted in the lowest workload measure for each subject and a workload index of 100 was assigned to the approach with the highest workload measure for the subject. The normalized workload index on approach i of subject j was then computed by

$$\text{Normalized WLX}_{ij} = \frac{\text{WLX}_{ij} - \min_i \{\text{WLX}_{ij}\}}{\max_i \{\text{WLX}_{ij}\} - \min_i \{\text{WLX}_{ij}\}} \times 100 \text{ percent} \quad (4)$$

Experimental Design

The experimental variables to be investigated in this study were the pilot's participation level in the piloting task, the workload induced by the control dynamics and by external disturbances, and the pilot's failure detection performance.

The experiment involved four levels of control participation:

- a. "Passive monitoring", with autopilot coupling in all axes, including autothrottle.
- b. "Yaw manual", with autopilot coupling in the pitch axis and autothrottle coupled.
- c. "Pitch manual", with autopilot coupling in the yaw axis only.
- d. "Fully manual".

There were three levels of wind disturbance:

- a. No wind.
- b. A 45° tailwind of 5 knots, gusting to 15 knots.
- c. A 45° tailwind of 10 knots, gusting to 30 knots.

Three failure conditions were used:

- a. No failure.
- b. Failure in the yaw axis. In this condition the autopilot, if coupled, or the flight director would steer the airplane away from the localizer course to intercept and track a course parallel to the nominal path but translated by a distance corresponding to one dot deviation (1.25°) at the point of failure occurrence. This resulted in a one-dot angular error about 100 seconds after the initiation of the failure. This type of failure was chosen, rather than a runaway failure, as it was quite subtle and therefore it provided a good measure of the limits of the pilot's failure detection capability.
- c. Failure in the pitch axis, which resulted in a one-dot deviation (0.35° of angular error) approximately 30 seconds after the occurrence of the failure.

Failures were presented only between the altitudes of 1800 and 800 feet; each approach was terminated either at touchdown or when a positive rate of climb has been established following the initiation of a go-around by the subject. The selection of the failure altitude was randomized, as was the selection of the direction of the failure (left-right in a lateral failure mode, up-down in a pitch failure mode). Workload levels and failure detection performance were investigated in separate experiments, to avoid possible contamination of failure detection data by the presence of a concomitant subsidiary task; the "no failure" condition was incorporated in the design so that the subjects

would not anticipate a failure on each and every approach.

RESULTS AND DISCUSSION

It seems clear from Figures 4 and 5 that the side-task scores were sensitive to variations both in the disturbance level and the participation mode. Indeed, analysis of variance under the hypothesis that the effects of the disturbance and of the participation were additive revealed that the variations in workload scores as a function of participation mode were significant at the $P \ll 0.01$ level and as a function of the severity of the disturbance - at the $P < 0.05$ level.

There was, however, no significant difference between workloads at the two low disturbance levels, namely, calm air and a quartering wind of five knots, gusting to fifteen knots. It was assumed, and it was verified by pilots' comments, that the components of the wind parallel and normal to the final approach path, 3.5 knots gusting to 10.6 knots, were not strong enough to induce workload significantly higher than that induced by piloting the simulated aircraft in calm air. Consequently, these two disturbance levels were combined in the analysis and the data were treated as if there were only two distinct disturbance levels, "low" and "high".

An additive model was used in the regression of workload scores on the disturbance levels and participation modes, to yield

$$WLX(P,D) = W_1(P) + W_2(D) \quad (5)$$

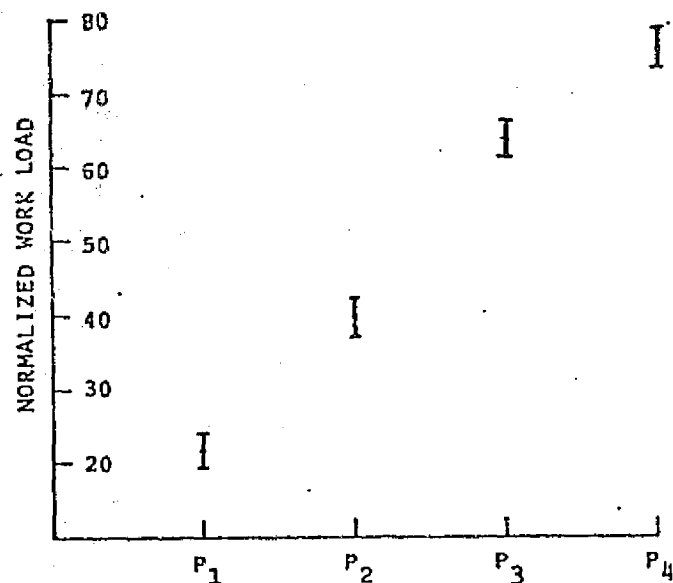
where WLX is the normalized workload score

P is the participation mode

D is the disturbance level

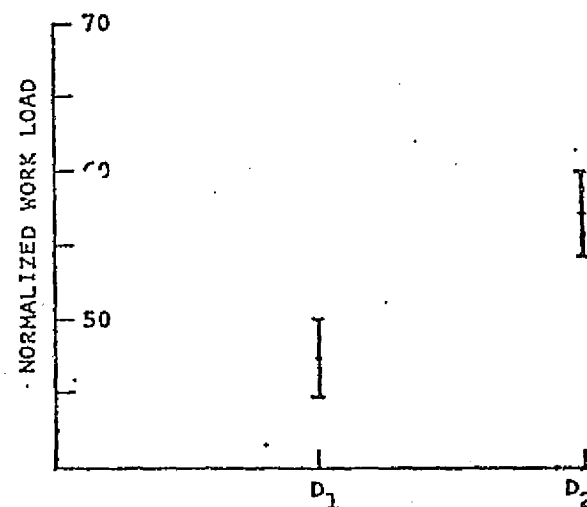
$$W_1(P) = \begin{array}{l} 18.7 \text{ for the fully-automatic mode} \\ 36.6 \text{ for split-axis, yaw manual mode} \\ 61.0 \text{ for split-axis, pitch manual mode} \\ 72.9 \text{ for the fully manual mode} \end{array}$$

$$\text{and } W_2(D) = \begin{array}{l} 0 \quad \text{for the "low" disturbance level} \\ 9.82 \text{ for the "high" disturbance level} \end{array}$$



Normalized Workload Index at Four
Participation Modes
P1 - Fully Automatic
P2 - Split Axis, Yaw Manual
P3 - Split Axis, Pitch Manual
P4 - Fully Manual

Figure 4.



Normalized Workload Index at Two
Disturbance Levels
D1 - Calm Air
D2 - 10 kt. Wind, Gusting to 30 kt.

Figure 5.

These values yielded workload-participation mode correlation significant at $P < 0.001$ and workload-disturbance correlation significant at $P < 0.05$.

Detection performance was analyzed in terms of detection time and accuracy. Detection time was defined as the elapsed time between the occurrence of a failure and the verbal report by the subject that the failure has been detected and identified. Accuracy was measured by the fraction of failures that were missed altogether. We differentiated between approaches in which a failure went unreported but which resulted in a successful touchdown and approaches in which a failure was missed and which did not terminate in a successful landing because of gross error in the failed axis. The latter are shown in Tables 1 and 2; the numbers in parentheses represent the fraction of all missed failures, whether or not they resulted in a successful landing.

In all, 90 approaches were flown in which a longitudinal failure occurred; of these, 8 went unreported, 6 of which did not terminate in a successful landing. Of the 90 lateral failures presented, 9 were missed; of these, 6 did not terminate in a successful landing.

A very interesting pattern is obvious from Tables 1 and 2 and from Figures 6 and 7: All failures in an automatically-controlled axis were detected in consistently short times; between 9 and 17 percent of the failures which occurred in a manually-controlled axis were not detected at all, and the ones that were required considerably longer detection times. The difference between the mean detection times in an automatic and manual mode was highly significant at the $P < 0.01$ level.

We hypothesized that this difference in detection performance was due, in part, to the increased involvement of the pilot in the control task in the manual mode and, in part, to the increased workload levels associated with manual control; we set out to separate the individual effects of these factors on the failure detection performance.

In Figures 6 and 7 the mean detection times of pitch and yaw failures, respectively, are plotted as functions of the corresponding mean workload levels for the four participation modes. The following relationships are evident:

1. Detection times in a manually-controlled axis are longer than detection

TABLE 1
Fraction of Missed Longitudinal Failures
in Percent

Participation Mode	disturbance Level			Overall
	1	2	3	
Monitor	0.	0.	0.	0.
Control Yaw	0.	0.	0.	0.
Control Pitch	12.5 (12.5)	0. (14.3)	12.5 (12.5)	8.7 (13.0)
Manual Control	0. (12.5)	14.3 (14.3)	37.5 (37.5)	17.4 (21.7)
Overall	3.3 (6.7)	3.3 (6.7)	13.3 (13.3)	6.7 (8.9)

TABLE 2
Fraction of Missed Lateral Failures
in Percent

Participation Mode	Disturbance Level			Overall
	1	2	3	
Monitor	0.	0.	0.	0.
Control Yaw	25.0 (37.5)	14.3 (14.3)	12.5 (37.5)	17.4 (30.4)
Control Pitch	0.	0.	0.	0.
Manual Control	14.3 (14.3)	0.	14.3 (14.3)	9.1 (9.1)
Overall	10.0 (13.3)	3.3 (3.3)	6.7 (13.3)	6.7 (10.0)

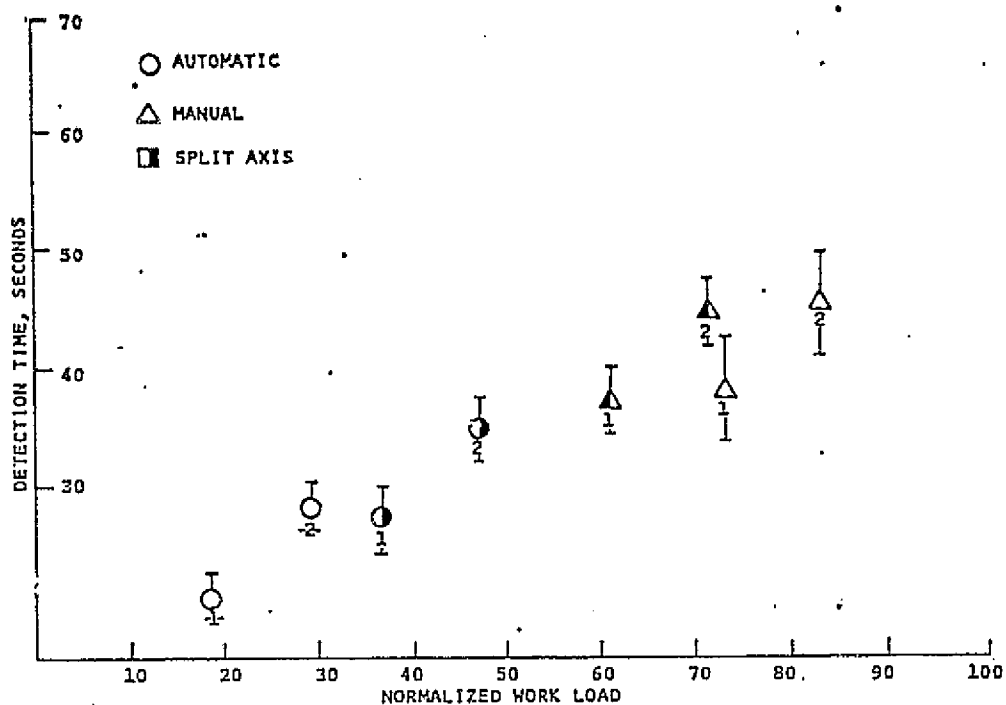


Figure 6. Mean Longitudinal Detection Times

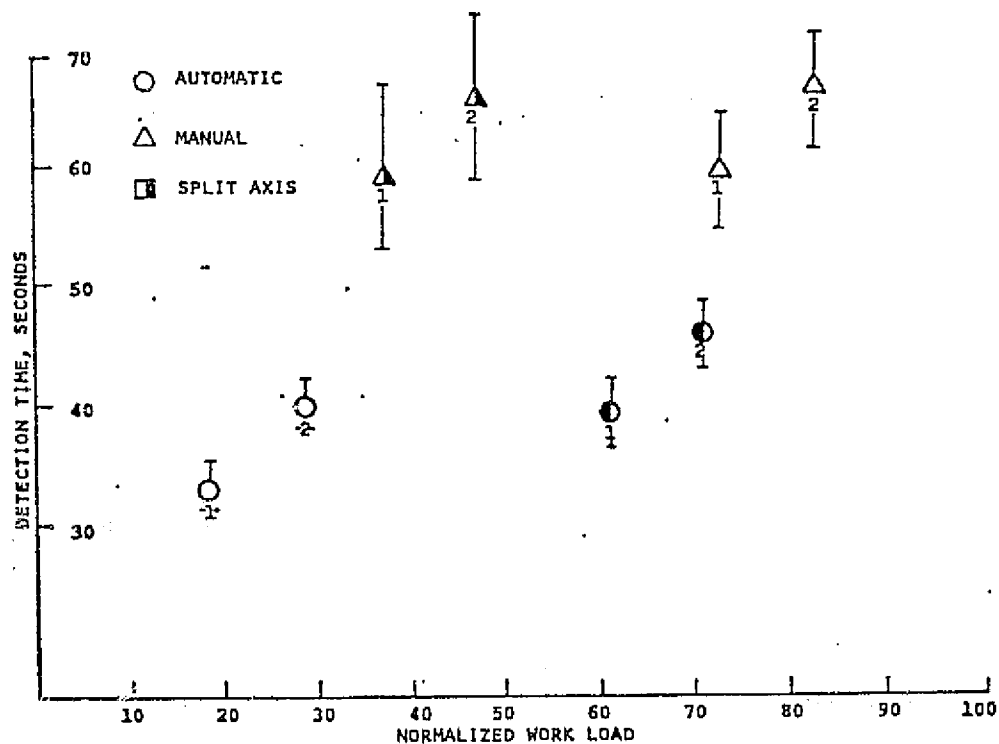


Figure 7. Mean Lateral Detection Times

- times in an automatically-controlled axis.
- 2. Detection times for lateral failures are significantly longer than detection times for longitudinal failures at comparable workload levels
- 3. Detection times increase in direct relationship to workload ($\rho = 0.322$ for $n=163$ pairs).

We assumed that the failure detection mechanism of the human operator acts similarly in both lateral and longitudinal axes; any difference in performance between these axes is due to differences in the plant dynamics and in display variables only, not to differences in processes internal to the operator. This assumption of equivalence between the lateral and longitudinal axes has been made, either explicitly or implicitly, by many investigators. It is based on the theory that the human behaves optimally with respect to his task (cf. Smallwood, 1967) in all axes, and that the operator adjusts his describing function to match the task (Young, 1969).

Longitudinal and lateral failure detection data were thus pooled; detection times were regressed on the type of failure (longitudinal or lateral) and on the control mode in the failed axis, with the workload index as a covariate, based on the following additive model:

$$T_{\text{detection}} = T_0 + \alpha(\text{control mode}) + \beta(\text{failed axis}) + \gamma(\text{workload}) \quad (6)$$

A solution was obtained for the regression coefficients α , β and γ :

$$T_{\text{detection}} = 20.9 + 16.5 \underline{M} + 15.4 \underline{A} + 0.10 \text{ WLX} \quad (7)$$

where $\underline{M} =$ 1 if the failed axis is controlled manually

$\underline{M} =$ 0 otherwise

$\underline{A} =$ 1 if the failure occurs in the lateral axis

$\underline{A} =$ 0 if the failure occurs in the longitudinal axis

WLX = the normalized workload index

and $T_{\text{detection}}$ is measured in seconds.

The relationship is plotted in Figures 8 and 9 for longitudinal and lateral

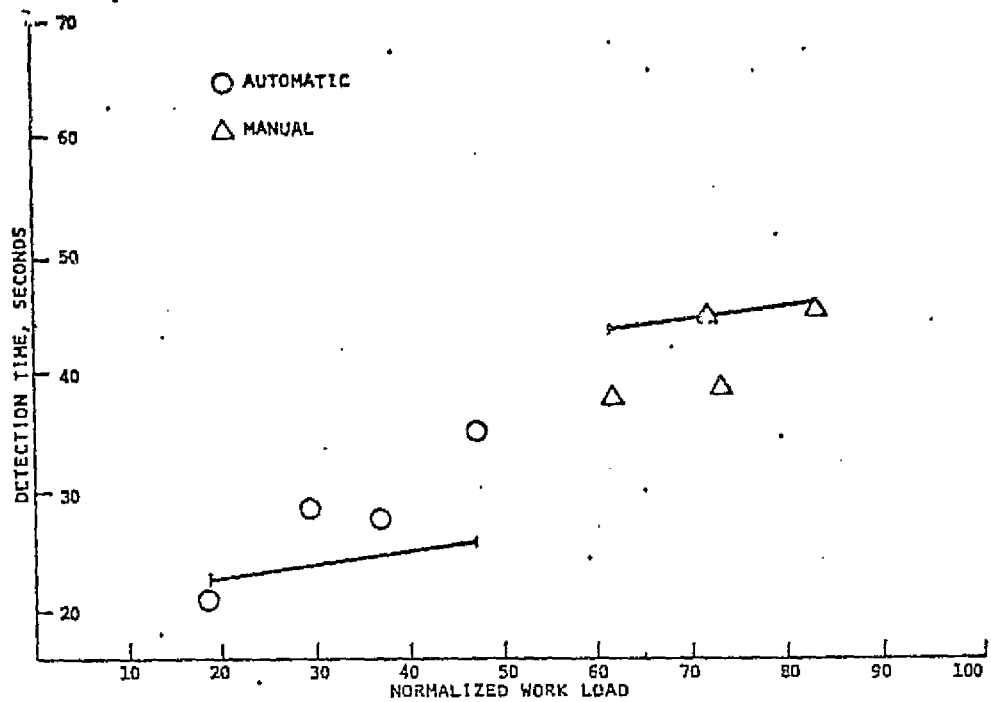


Figure 8. Predicted Longitudinal Detection Times

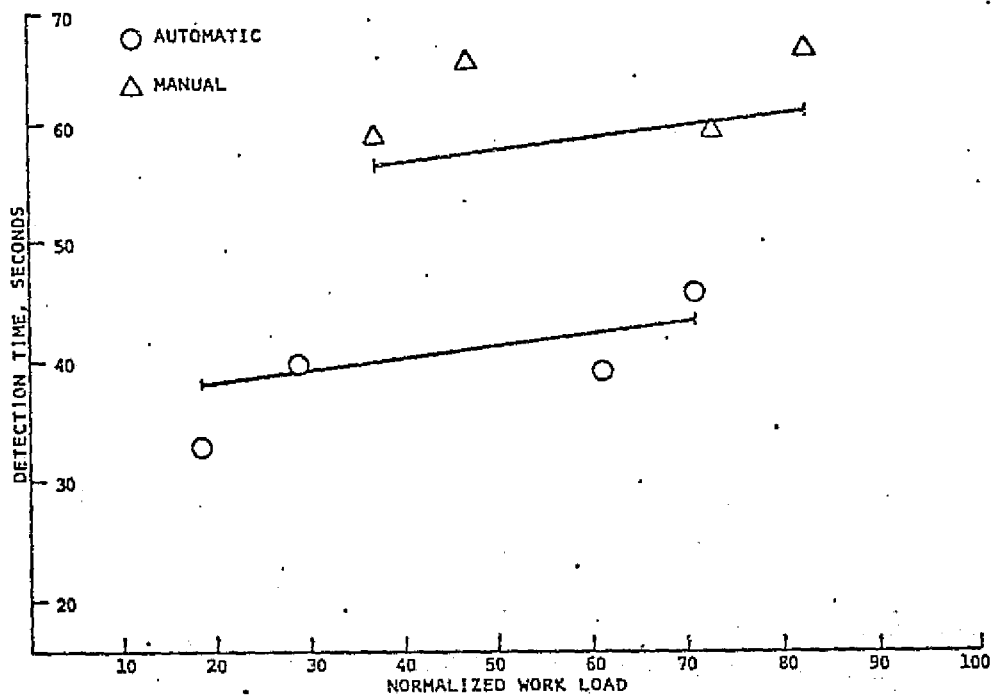


Figure 9. Predicted Lateral Detection Times

failures, respectively. Mean detection times at the corresponding mean workload levels are also shown for comparison. The model correlates well with the data, with $p_{n=163} = 0.531$, significant at the $P \ll 0.001$ level.

CONCLUDING REMARKS

Our goal in this research was to identify the participation mode and workload level which optimize the pilot's failure detection performance; this subject is treated in considerably more detail elsewhere (Ephrath, 1975).

Our results indicate quite clearly that a coupled, fully-automatic landing with the lowest possible workload is called for in Category III operations, with the crew monitoring the progress of the approach via cockpit displays: Failure-detection performance in all other control modes was unacceptable for commercial operations. Performance monitors and fault annunciators may alleviate the problem somewhat but they are inadequate at altitudes below 100 feet (Vreuls *et al*, 1968); also, they are not infallible, and additional warning lights and buzzers in the cockpit provide more opportunities for malfunctions and for crew confusion.

REFERENCES

- Ekstrom, P.J., *Analysis of Pilot Workloads in Flight Control Systems with Different Degrees of Automation*, IRE International Congress on Human Factors Engineering in Electronics, Calif., May 1962.
- Ephrath, A.R., *Pilot Performance in Zero Visibility Precision Approach*, Ph.D. Thesis, Massachusetts Institute of Technology, May 1975.
- Gai, E.G. and R.E. Curry, *Failure Detection by Pilots During Automatic Landings: Model and Experiments*, Annual Conference on Manual Control, California, May 1975.
- Gainer, C.A. *et al*, *All Weather Landing Simulation for Category III Airborne*

Configuration: Flight Directors and Split Axis Control, Bunker-Ramo Corp. California, SRDS-RD-67-56/1, July 1967.

Levison, W.H., *A Model for Task Interference*, Annual Conference on Manual Control, Ohio, April 1970.

Levison, W.H., *A Control-Theory Model for Human Decision Making*, Annual Conference on Manual Control, California, June 1971.

Phatak, A.V. and G.A. Bekey, *Decision Processes in the Adaptive Behavior of Human Controllers*, IEEE Transactions on Systems Science and Cybernetics, V. SSC-5, 4:339-351, October 1969.

Schrenk, L.P., *Aiding the Decision Maker - A Decision Process Model*, IEEE Transactions on Man Machine Systems, V. MMS-10, 4:204-218, Dec. 1969.

Smallwood, R.D., *Internal Models and the Human Instrument Monitor*, IEEE Transactions on Human Factors in Electronics, V. HFE-8, 3:181-187, September 1967.

Spyker, D.A. et al, *Development of Techniques for Measuring Pilot Workload*, NASA CR-1888, 1971.

Vreuls, D. et al, *Pilot Failure Detection Performance with Three Levels of Fault Warning Information*, Bunker-Ramo Corp., California, SRDS-RD-68-9, February 1968.

Wewerinke, P.H., *Human Operator Workload for Various Control Situations*, National Aerospace Laboratory, N.L.R., Amsterdam, the Netherlands.

Young, L.R. et al, *The Adaptive Dynamic Response Characteristics of the Human Operator in Simple Manual Control*, NASA TN D-2255, 1964.

Young, L.R., *On Adaptive Manual Control*, IEEE Transactions on Man Machine Systems, V. MMS-10, 4:292-331, December 1969.

APPENDIX B

PRECEDING PAGE BLANK NOT FILMED

RESEARCH NOTE:

CHANGES IN PILOT WORKLOAD DURING AN INSTRUMENT LANDING

Arye R. Ephrath
Renwick E. Curry

Man Vehicle Laboratory
Massachusetts Institute of Technology

This research was sponsored by NASA Grant NGR 22-009-733.

Introduct

It has been recognized that when the role of the pilot changes from that of an active controller to that of a monitor of an automatic system corresponding changes take place in his workload level (Ekstrom, 1962). However, in pilot performance studies to date these effects were averaged over the entire range of a particular flight phase - usually the landing approach. Useful information may be lost by such averaging. Pilot's performance has been shown to vary greatly among different segments of an approach while flying by reference to instruments (DeCelles et al, 1970; Gainer et al, 1967; Ephrath, 1975). Such variations in performance may be caused by corresponding variations in the pilot's instantaneous workload level during the landing approach.

If indeed such is the case, and the observed deterioration in the pilot's instrument-flying performance is caused by excessive workload levels, then it is evident that improved piloting performance may be achieved through cockpit designs that eliminate the causes of these excessive workloads. The purpose of our work

is hence threefold: to examine the effect on workload of different levels of automation, to document the variations in the instantaneous workload level of a pilot during an instrument approach for a given level of automation; and to determine the particular phases of the approach during which the workload levels are highest as a prerequisite to identifying their causes.

Method

In the course of this study, fifteen professional airline pilots flew a total of 180 instrument landing approaches in a fixed-base cockpit simulator which was programmed to duplicate the dynamics of a large jet transport aircraft; the entire experiment was controlled on-line by an ADAGE AGT/30 digital computer.

The pilots instantaneous workload level was measured by means of a non-loading warning-lights-type subsidiary task using two small (1/8" diameter) red lights mounted above each other outside the pilot's peripheral field of vision at 75° to the right of the center of the instrument panel. At random times, during the approach, the upper or lower light was lit with equal probability. A correct response from the subject consisted of a proper motion of a control-yoke-mounted, three-position switch; that is, pushing the switch up or down in response to the top or bottom light respectively. A correct response by the subject caused the light to turn off; a "hit"

was counted and the subject's response time recorded. In the absence of a correct response the light remained on for two seconds, then was turned off, and a "miss" was counted. An incorrect response by the subject (that is, pushing the switch the wrong way) was also counted as a "miss".

After the light was turned off, a time delay followed, uniformly distributed between 0.5 and 5.0 seconds, and the process was repeated. The workload index, WLX, was computed by

$$WLX = 100 \frac{\alpha(\text{cumulative response time}) + \beta(\text{number of "misses"})}{(2\alpha + \beta)(\text{number of "hits"} + \text{number of "misses"})}$$

which resulted in a workload-index value between 0 and 100. The coefficients α and β were chosen to maximize the correlation between this index and physiological measures of workload (Spyker et al, 1971).

A more detailed discussion of this subsidiary task and of the design of the simulator system can be found elsewhere (Ephrath, 1975).

Each subject flew twelve simulated approaches from a point well beyond the outer marker to touchdown at four levels of automation or participation: fully manual, split-axis with manual yaw (pitch and power being controlled by an autopilot), split-axis with manual pitch (yaw being controlled by an autopilot), and fully automatic, with the pilot monitoring the instruments. Three approaches were flown by the subjects at each automation level, with the order of presentation randomized to eliminate learning effects. To this end, each pilot was also allowed ample practice in the simulator prior to commencement of the formal experiment.

Results

Workload data were segregated by participation modes; mean scores were extracted at 300 foot altitude intervals between the altitudes of 2000 and 500 feet, and at 100 foot intervals between 500 feet and touchdown. The results, averaged over all pilots, are shown in Figures 1 through 4.

Insert Figures Here

Figures 1 through 4 reveal the asymptotic behavior of the pilot's workload as inferred from the index. The workload index is essentially constant at altitudes above approximately 500 feet, but there is a very marked increase in the workload at lower altitudes where the workload index appears to be inversely related to distance-to-go. This increase may be at least partially due to the non-linear increase in the display sensitivity. As the standard flight instruments display angular rather than linear deviations from the localizer course and from the glide path, the display increases in sensitivity as the distance to touchdown decreases. However, this increase in pilot workload was observed even when the simulated aircraft was controlled by the autopilot and the subject acted merely as a monitoring element; under autopilot control, the indicator instruments barely deviated from the zero-error position and, therefore, any changes in their sensitivity could not affect the monitoring subject's workload. This seems to suggest the effects

of other factors such as, possibly, changes in the pilots' mental state arising from their awareness of the proximity of the ground and of the associated increase in the penalty for error.

The observed increase in pilot's workload at low altitudes is quite interesting, as it is precisely this phase of a landing during which the pilot's instrument-flying performance becomes unacceptable and a transition to visual flight is mandatory; if visual reference with the ground cannot be established, the approach must be aborted (DeCelles et al, 1970).

It is conceivable that display design might reduce the additional workloads imposed on the pilot as a result of increased display sensitivity; further studies are necessary however, to determine whether other factors are inherent in this situation and whether an improvement is possible in the pilot's workload - and his flying performance - during the last phase of the landing approach.

5

REFERENCES

- DeCelles, J.L. et al, "The fail safe landing", Report of the ALPA All-Weather Flying Committee, ALPA's 17th Air Safety Forum, San Francisco, CA, July 1970.
- Ekstrom, P.J. "Analysis of pilot workload in flight control systems with different degrees of automation", IRE International Congress on Human Factors Engineering in Electronics, Long Beach, CA, May 1962.
- Ephrath, A.R., "Pilot performance in zero-visibility precision approach", Ph.D. Thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, June 1975.
- Gainer, C.A. et al, "Summary of studies on flight directors and split axis control", Bunker Ramo Corporation, Canoga Park, CA, July 1967, SRDS-RD-67-56/1.
- Spyker, D.A. et al, "Development of techniques for measuring pilot workload", Honeywell, Inc. Roseville, Minn. November 1971, NASA CR-1888.

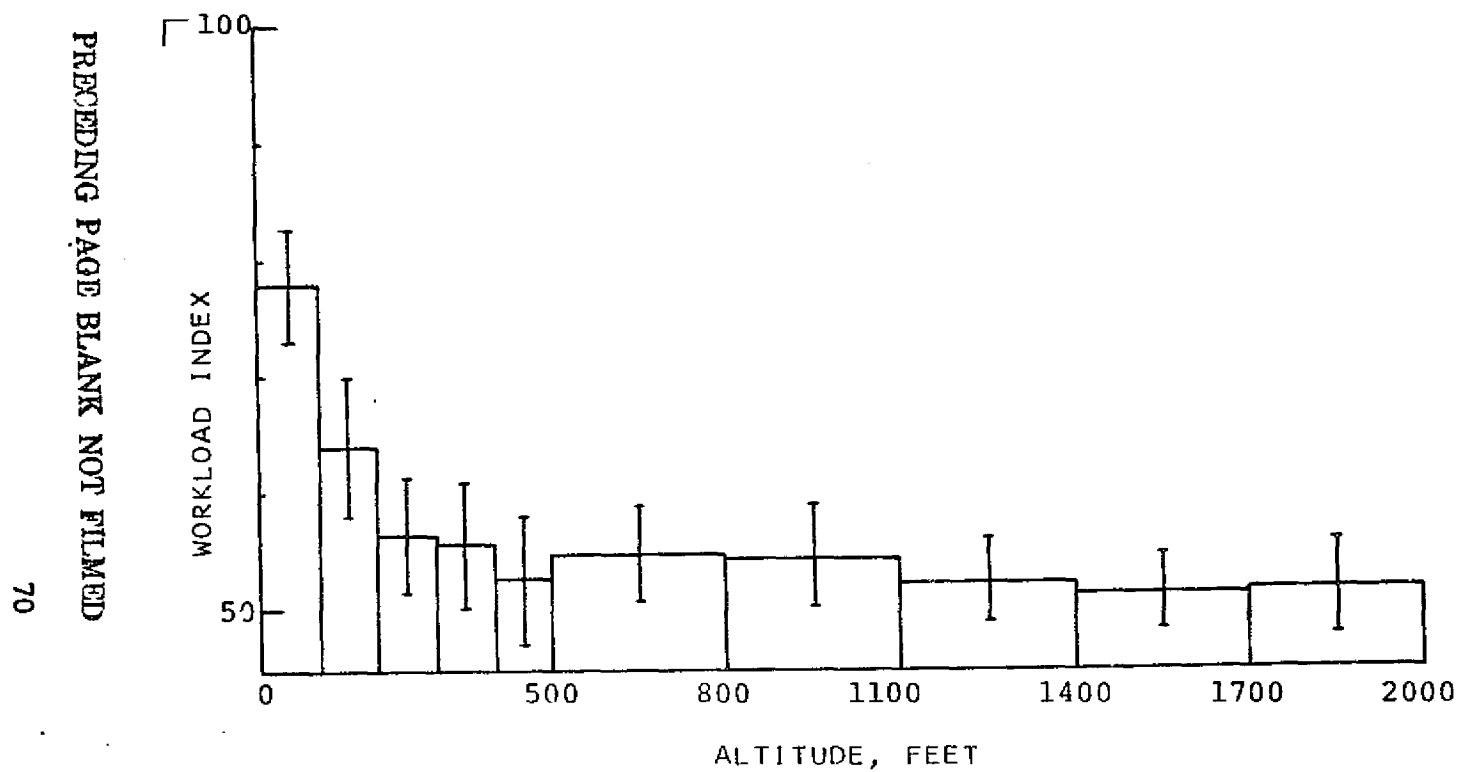


Fig. 3.12 Non-Normalized Workload Scores vs. Altitude - Monitoring

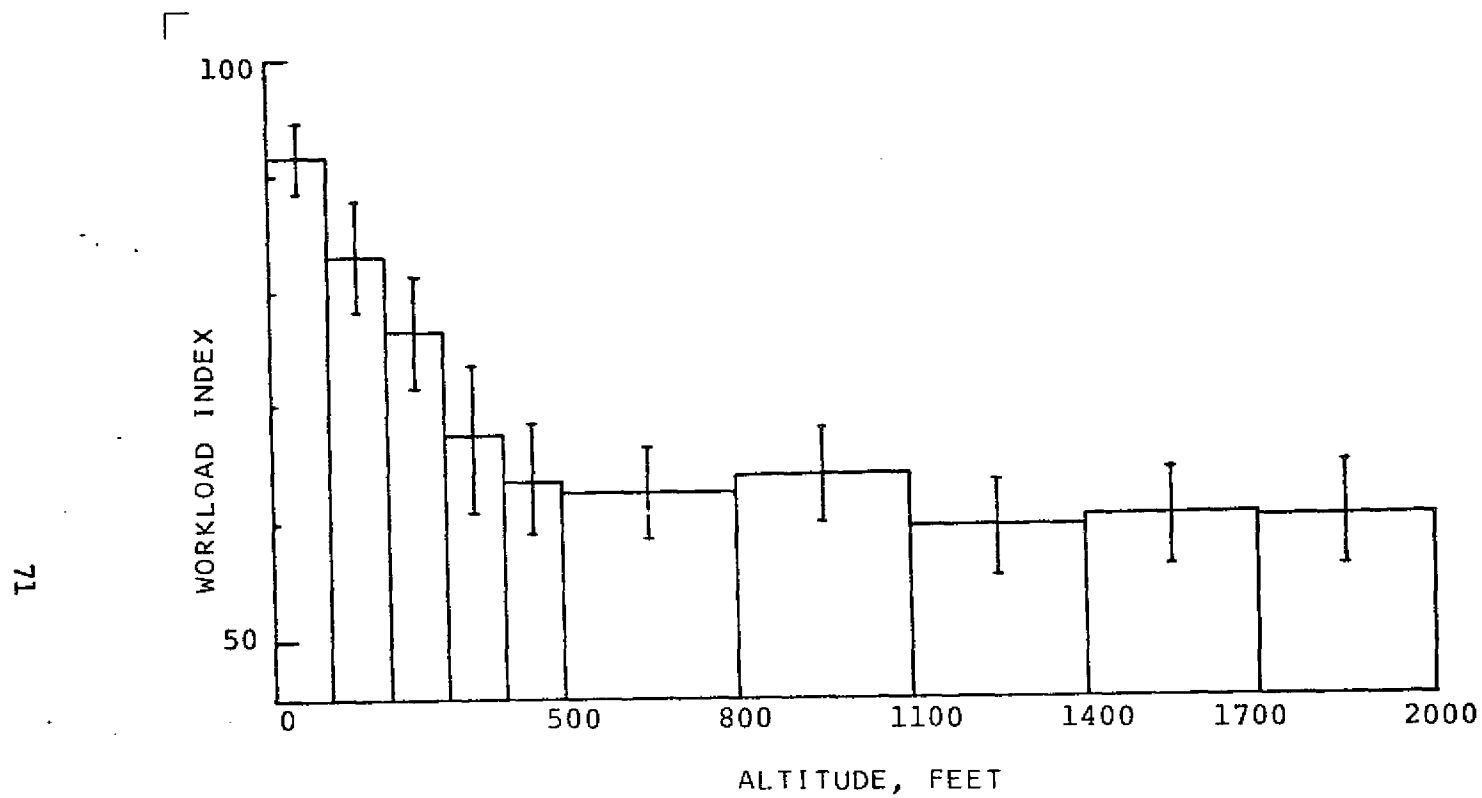


Fig. 3.13 Non-Normalized Workload Score vs. Altitude -
Lateral Control

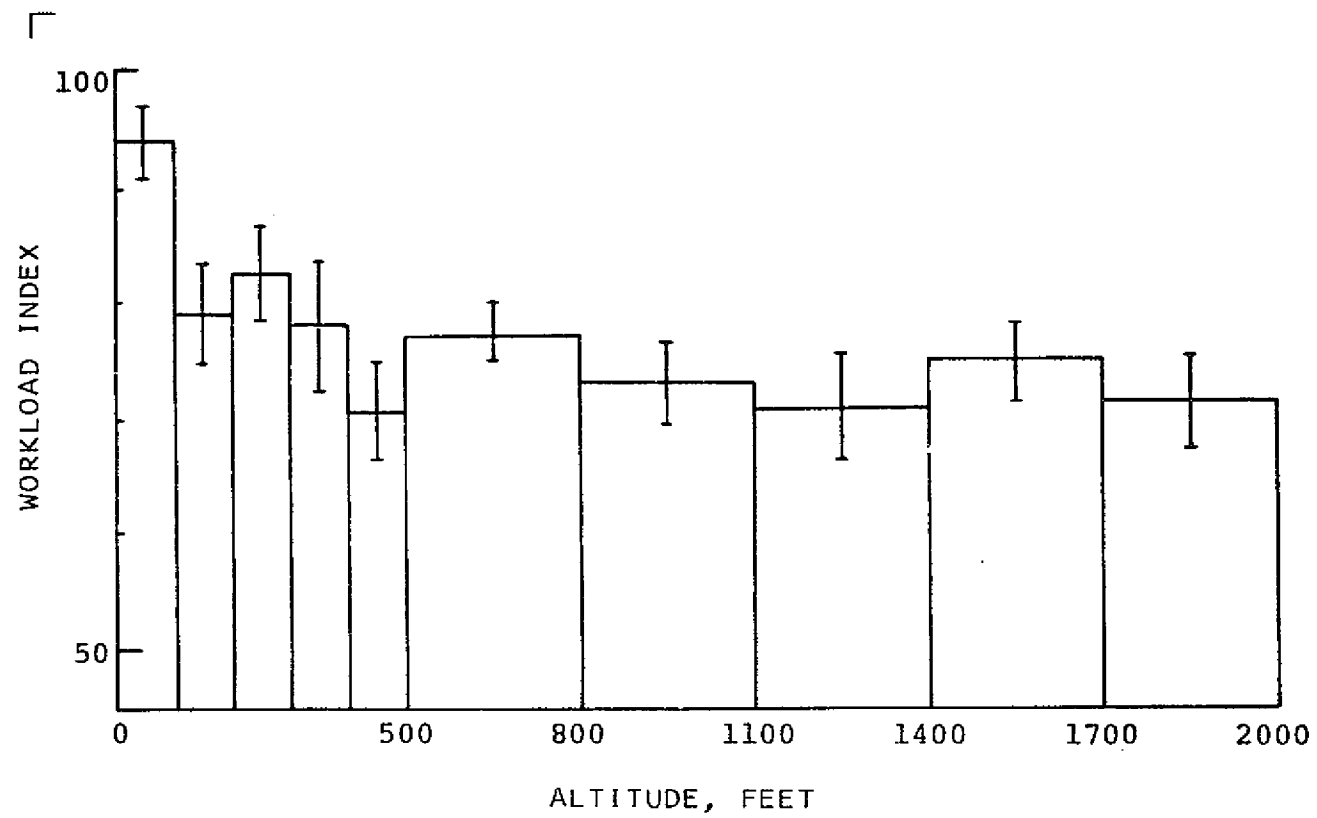


Fig. 3.14 Non-Normalized Workload Scores vs. Altitude -
Longitudinal Control

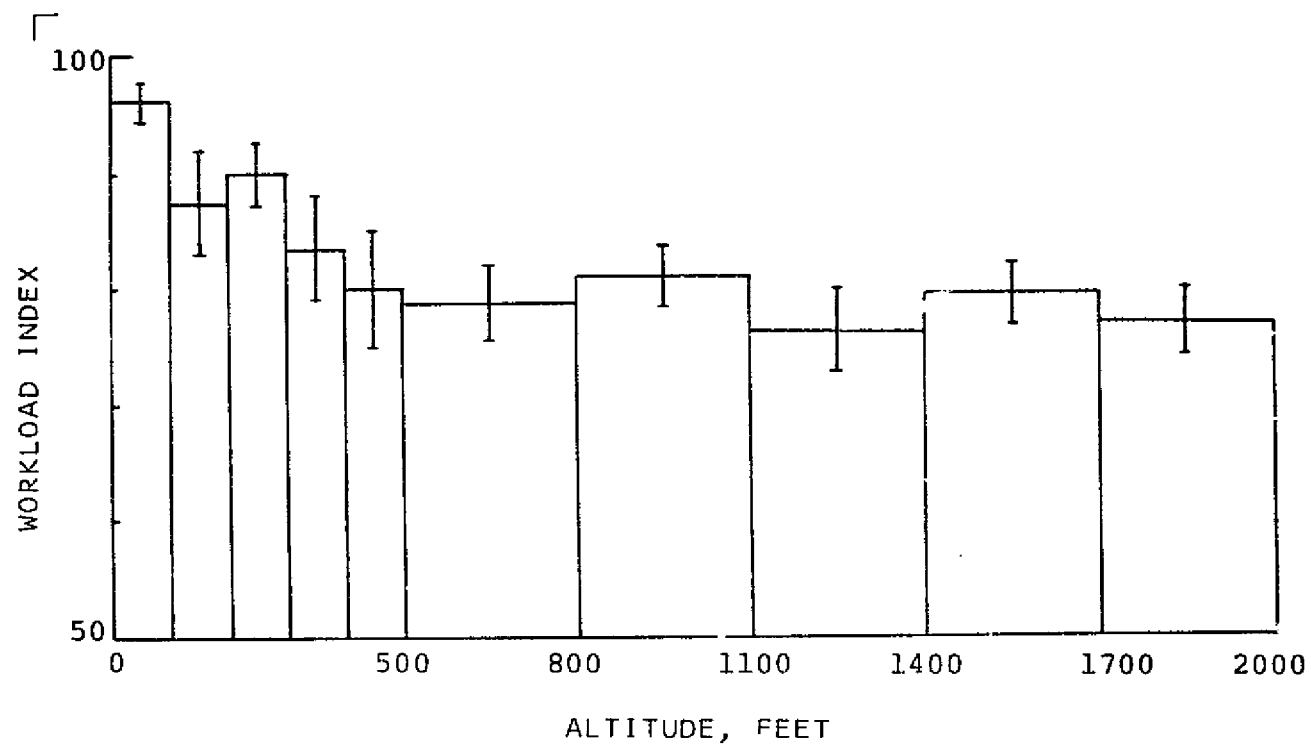


Fig. 3.15 Non-Normalized Workload Scores vs. Altitude -
Manual Control

APPENDIX C

PRECEDING PAGE BLANK NOT FILMED

FAILURE DETECTION BY PILOTS DURING AUTOMATIC LANDING: MODELS AND EXPERIMENTS *

Eli G. Gai
C.S. Draper Laboratories
Room DL7-173
75 Cambridge Parkway
Cambridge, MA 02142

R.E. Curry
Man Vehicle Laboratory
Room 37-219
M. I. T.
Cambridge, MA 02139

ABSTRACT

A model of the pilot as a monitor of instrument failures during automatic landing is proposed. The failure detection model of the pilot consists of two stages: a linear estimator (Kalman Filter) and a decision mechanism which is based on sequential analysis. The filter equations are derived from a simplified version of the linearized dynamics of the airplane and the control loop. The perceptual observation noise is modelled to include the effects of the partition of attention among the several instruments. The final result is a simple model consisting of a high pass filter to produce the observation residuals, and a decision function which is a pure integration of the residuals minus a bias term.

The dynamics of a Boeing 707 were used to simulate the fully coupled final approach in a fixed base simulator which also included failures in the airspeed, glideslope, and localizer indicators. Subjects monitored the approaches and detected the failures; their performance was compared with the predictions of the model with good agreement between the experimental data and the model.

INTRODUCTION

The introduction of the "all weather" automatic landing system changes the role of the pilot during landing. Under normal conditions, the pilot is not in the control loop, but his main task is to monitor the proper operation of the automatic system. This, of course, shifts his role from manual controller to decision maker.

The problem of modelling the pilot as a controller has been addressed by several researchers, and satisfactory models exist using classical control theory (1) or optimal control theory (2). Models for the pilot as a failure detector have only recently been addressed (3), and some conjecture has been suggested (4).

In this paper, a model of a pilot as a monitor of system failures is proposed and applied to an automatic landing. The model consists of two stages: a linear estimator and a decision mechanism. The linear estimator is the Kalman Filter which is similar to that used in the optimal controller model; the outputs used here are the measurement residuals rather than the estimates. The decision mechanism is based on sequential analysis (5) modified for the special case of failure detection (6). Experiments were designed to validate the proposed model in which subjects monitored failures in simulated automatic ILS

*Sponsored by NASA Grant NGR 22-009-733, NASA Ames Research Center

approaches in a fixed-based jet transport cockpit. The results of these experiments are then compared to the prediction of the model.

PROBLEM STATEMENT AND SIMPLIFICATION

A functional block diagram of the failure detection model is shown in Figure 1. A basic assumption in the structure of the first stage (the estimation) is that the dynamical characteristics of the system that produces the input signals are known by the observer. Therefore, before the modelling of the failure detection system, we will discuss the model that the pilot has in mind (the internal model) for the airplane dynamics and control loops.

The true airplane dynamics, when angular accelerations are neglected, can be defined by nine first order nonlinear differential equations. Two decoupled autopilots are used to regulate the vertical error between the aircraft position and the glideslope beam, and the horizontal (angular) error between aircraft position and localizer beam. In addition, a third control loop controls the aircraft airspeed. This configuration was used in the simulation that automatically landed the Boeing 707 dynamics used in our experiments (7).

Since the pilot is outside the control loop his inputs consist only of the displayed variables on his instrument panel. If the control system is designed properly, these displays will show nominal values with variation due to outside perturbations. It seems reasonable to assume that the pilot will use the linearized version of the automatic system around the nominal values. In addition, we will assume that the longitudinal and lateral dynamics are decoupled in the pilot's model. The block diagrams of the three control loops are shown in Figure 2, and the basic configuration was taken from reference 8. Therefore the three closed loop transfer functions are given by

$$\frac{\delta u}{\delta u_n} = \frac{10(S+0.1)}{(S+8.8)(S+.98)(S+0.13)} \quad (1)$$

where: u - velocity
 γ - flight path angle
 θ - pitch
 ψ - heading

$$\frac{\delta \gamma}{\delta \theta_n} = \frac{31.3}{(S+0.5)(S+5.5)(S^2+5.45S+11.4)} \quad (2)$$

$$\frac{\delta \psi}{\delta \psi_n} = \frac{47}{(S^2+11S+58)(S^2+1.5S+0.81)} \quad (3)$$

The letter δ is used to identify the inputs as perturbations rather than commands, and the outputs are the responses to these perturbations. The subscript n is used because the input perturbations are modelled as zero mean white Gaussian processes. The source for these perturbations is usually the wind gusts, and therefore the inputs to the subsystem are correlated and are derived from the amplitude and direction of these gusts. Two of the above subsystems are of fourth order and one is of third order. Another integration of each subsystem output is needed to obtain the aircraft position. Therefore, we assume that the pilot uses only the dominant poles, namely the ones

with the longer time constants. The final model that is used in the pilot model is

$$\frac{\delta u}{\delta u_n} = 1/(s+1) \quad (4)$$

$$\frac{\delta \gamma}{\delta \theta_n} = 1/((s+0.5)(s+2.7)) \quad (5)$$

$$\frac{\delta \psi}{\delta \psi_n} = 1/(s^2+1.5s+0.81) \quad (6)$$

Note that the steady state gain and the steady state variance of the response to a stationary random input were kept the same as in the original systems.

Having three decoupled systems, we can define 8 state variables by transforming equations (4) to (6) to their state space format. Define

$$\begin{aligned} x_1 &= \int \delta u & x_2 &= \delta u & x_3 &= \int \delta \gamma & x_4 &= \delta \gamma \\ x_5 &= 0.585 \delta \alpha & x_6 &= \int \delta \psi & x_7 &= \delta \psi & x_8 &= g \delta \phi / V_0 \end{aligned} \quad (7)$$

The dynamics in matrix notation are:

$$\dot{\underline{X}} = \underline{F}\underline{X} + \underline{G}\underline{U} \quad (8)$$

where

$$\underline{F} = \begin{bmatrix} F_1 & 0 & 0 \\ 0 & F_2 & 0 \\ 0 & 0 & F_3 \end{bmatrix} \quad F_1 = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \quad F_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -1.35 & -3.2 \end{bmatrix}$$

$$F_3 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -0.81 & -1.5 \end{bmatrix} \quad \underline{U}^T = [\delta u_n, \delta \theta_n, \delta \psi_n] \quad \underline{G}^T = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.35 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.81 \end{bmatrix}$$

The perturbations in the aircraft position in terms of the above state variables are given by (using the fact that γ_0 is small):

$$\begin{aligned} \delta x &= \cos \psi_0 x_1 - v_0 \sin \psi_0 x_6 \\ \delta y &= \sin \psi_0 x_1 - v_0 \cos \psi_0 x_6 \\ \delta z &= \gamma_0 x_1 + v_0 x_3 \end{aligned} \quad (9)$$

where the subscript 0 designates nominal values.

All the variables that are displayed to the pilot can now be represented as linear functions of the state variables

1. Glideslope Indicator: $y_1 = (-\cos \psi_0 / x_N^2 + \gamma_0 / x_N) x_1 + v_0 x_3 / x_N + v_0 \sin \psi_0 x_6 / x_N^2$
2. Localizer: $y_2 = [\cos \psi_0 / (1.23 - x_N)^2 + \sin \psi_0 / (1.23 - x_N)] x_1 + [v_0 \cos \psi_0 / (1.23 - x_N) - v_0 \sin \psi_0 / (1.23 - x_N)^2] x_6$
3. Airspeed indicator: $y_6 = x_2$

4. Altitude indicator: $y_4 = x_4 + x_5/0.585$ (pitch)
 $y_4 = v_0 x_8/g$ (roll angle)
5. Horizontal situation display: $y_6 = x_7$
6. Altimeter: $y_7 = \gamma_0 x_1 + v_0 x_3$
7. Vertical speed indicator: $y_8 = \gamma_0 x_2 + v_0 x_4$

x_0 in the above equations is the nominal value in the x direction which is time varying.

THE FAILURE DETECTION MODEL

In the last section we suggested a simplified linear model which the pilot is assumed to use as a model for the instrument output dynamics. On the basis of these dynamics the failure detection model that was suggested by Gai (6) is used. As mentioned before, the model consists of two stages: a linear estimator and a decision mechanism. The linear estimator, the Kalman Filter, is shown in Figure 3. It should be noted that due to the time dependency of the measurements the Kalman gain $K(t)$ is time varying. The filter produces estimates for the state $\hat{x}(t)$ and the measurements $\hat{y}(t)$ as well as the measurement error (residual) $\varepsilon(t)$ which is defined as

$$\varepsilon(t) = y(t) - \hat{y}(t) \quad (10)$$

Any of these three quantities can be used as an input to the decision mechanism. The observation residual is preferred for the following reasons:

1. The state variables are non-unique variables that can be defined in many ways, while the observation residual is unique and well-defined for the subject.
2. The dimension of the state is in general larger than the dimension of the residual.
3. The residual is more sensitive than the observation to the effect of the failure (9).
4. The residual is a zero mean white process (10) in the unfailed mode so successive observations are independent for Gaussian processes.

In the instrument failure mode, we will assume that a deterministic mean is added to the measurement so that the residual is still white Gaussian with the same variance but with a non-zero mean.

Since the pilot is using 3 instruments the problem of sharing of attention has to be accounted for. This is done through the measurement noise in the observer model (11). If the pilot is observing more than one instrument, his observation noise for each observation is increased by a constant factor that is inversely proportional to the fraction of attention that he spends monitoring that specific instrument. Finally, it should be noted that although the state equations are decoupled, the Kalman Filter is a coupled 8 dimensional system because of the coupling through the measurements. The model of the estimation scheme is shown in Figure 4.

The decision mechanism is based on sequential analysis (5). The classical sequential analysis uses the likelihood ratio $\lambda(m)$ as a decision function after m observations. Two criteria levels, A and B, are chosen, and the decision rule is

if $\lambda(m) \geq A$ choose "failure"
 if $\lambda(m) \leq B$ choose "normal"
 if $B < \lambda(m) < A$ take another observation

A and B are determined by the desired probability of false alarm $P(FA)$ and the probability of miss $P(MS)$ follows (5)

$$A = (1-P(MS))/P(FA) \qquad B = P(MS)/(1-P(FA)) \qquad (11)$$

Since in our case, both distributions are white Gaussian with equal variances and means zero and θ_1 (failure), the decision function (for $\theta_1 > 0$) is (6)

$$\lambda(m) = \sum_{i=1}^m \{\epsilon_i - \theta_1/2\} \qquad (12)$$

the upper criterion level is $(\ln A)/\theta_1$ (13)

and the lower criterion level is $(\ln B)/\theta_1$ (14)

The classical theory cannot be applied directly to the failure detection problem because the basic assumption in the derivation was that the same mode exists during the entire period. A failure detection problem is characterized by a transition from the normal mode to the failure mode at some random time t_f . In order to overcome this difficulty, we followed Chein (12) by:

1. Resetting the decision function to zero whenever $\lambda(m)$ is negative.
2. Using only an upper criterion level A_1 which is modified to keep the same mean time between two false alarms as before including the resetting.

The value of A_1 is related to A and B in equation (11) by the equation

$$A_1 - \ln A_1 - 1 = -\{\ln A + (A - 1)\ln B/(1-B)\}$$

The modified decision function is shown in Figure 5 and the block diagram of the basic decision mechanism is shown in Figure 6. For the case $\theta_1 < 0$, the decision function is

$$\lambda(m) = \sum_{i=1}^m (\epsilon_i + \theta_1/2) \qquad (15)$$

and only the lower criterion level is used. This criterion level is

$$-(\ln A)/\theta_1 \qquad (16)$$

The final block diagram of the decision mechanism is shown in Figure 7.

The operation of the proposed model is actually quite simple in principle. Its basic properties are:

1. A high pass filter as a first stage to obtain the residuals
2. Integration of the residual and comparison to a fixed threshold as a

decision mechanism (6).

3. Only three parameters control the performance of the model:
 - a. The parameter designating the mean of a "failed" process, θ_1
 - b. The signal to noise ratio of the observation noise in the Kalman Filter
 - c. The probabilities of the two types of error $P(FA)$ and $P(MS)$.

EXPERIMENTAL VALIDATION

Method

The Adage Model 30 Graphics Computer was used to simulate the non-linear dynamics of a Boeing 707 and the control loops (7). The output variables were fed into the instrument panel of a fixed based Boeing 707 simulator.

The simulation included the last five minutes of flight prior to touch down from 10 miles from threshold and at 2500 feet altitude; the approach and landing were fully automatic. The failures that were used were instrument failures, so that they affected only the output variables and were not fed back to the system. In order to limit the experimental requirements, we considered only failures in two instruments, the glideslope (GS) indicator and the airspeed (AS) indicator. Four levels of failures were included for each of the two instruments. All failures were deterministic step changes that were fed into the instrument through a single pole low pass filter with 0.1 second time constant. The magnitude of the failures for the AS indicator were

$$c_1 = 2\sigma_v \quad c_2 = 3\sigma_v \quad c_3 = 4\sigma_v \quad c_4 = 5\sigma_v \quad (17)$$

and for the GS indicator

$$c_1 = \sigma_{GS} \quad c_2 = 1.5\sigma_{GS} \quad c_3 = 2\sigma_{GS} \quad c_4 = 2.5\sigma_{GS} \quad (18)$$

where σ_v and σ_{GS} are the standard deviations of the perturbation from the nominal of the displayed variable on the AS and GS indicators respectively. The cutoff frequency of these perturbations were $\pi/6$ radians/sec. Two random number generators were used to choose the failure in each run; one determined the instrument and the other the size of the failure. In addition, a third random number generator was used to determine the time of failure, t_f .

There was a single failure in 90% of the runs. The high percentage of runs with failures was chosen to provide enough data in a reasonable experimental time. There was no feedback to the pilot concerning his performance because it was felt that such feedback would bias his decision, by driving him to try to compensate for previous errors, or to overrelax after several correct decisions.

Each subject participated in three experimental sessions each of which included 16 runs, or a total of 48 runs. When the pilot detected a failure, he pressed a button and the run was terminated. Otherwise, the run would go until touch down. After termination of each run the subject was asked to fill out a form in which he stated which instrument failed and how he detected the failure.

At the beginning of each session, a set of instructions was read to the subject. In particular, he was told that failure would either be in the AS or GS indicator, but that he could use other instruments for verification.

Results

The experimental results for two subjects are summarized in Table 1. The table shows the mean and standard deviation of the detection time for failures in the two instruments. The results are also shown in Figures 8 through 11. These figures include the mean detection times that were predicted by the model. These mean values were obtained by the use of a Monte Carlo simulation, with the same number of runs as in the experiment. The values for the three model parameters were:

$$\text{SNR} = 36 \quad P(\text{FA}) = P(\text{MS}) = 0.05 \quad \gamma_1 = 0.25 \quad (19)$$

The level of $P(\text{FA})$ was determined on the basis of the actual false alarm rate that was found in the experimental data. For both subjects, the predicted results seem to fit the experimental data well. Of course, better fit could be obtained by change of the parameters in equation (19).

CONCLUSIONS

In this paper we proposed a model for the performance of a pilot as a failure detector of instrument failures during an automatic landing. The model consists of two stages: a linear estimator and a decision mechanism. The linear estimator is the Kalman Filter which is determined from a simplified model of displayed-variable dynamics that are used by the pilot. The filter also accounts for the pilot's time sharing between instruments through the observation noise. The decision mechanism is based on classical sequential analysis with some modification for the failure detection case.

An experiment designed to test the validity of the model is described. In this experiment, subjects had to detect failures in the glideslope and air-speed indicators during a simulated landing in a Boeing 707 fixed base simulator. The results show that the predicted detection times fit the experimental data well. It should be remembered, however, that only instrument failures in the form of a change in the mean were discussed. Additional work is needed to verify the model to include failures that involve dynamic changes as well.

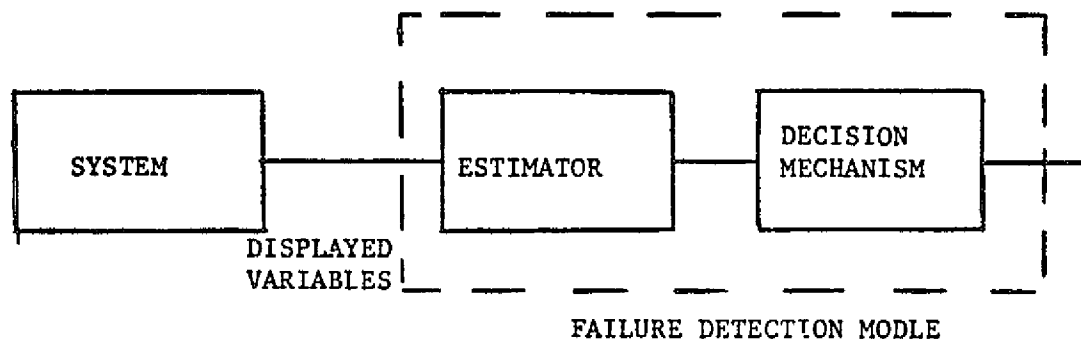


FIGURE 1 SCHEMATIC BLOCK DIAGRAM

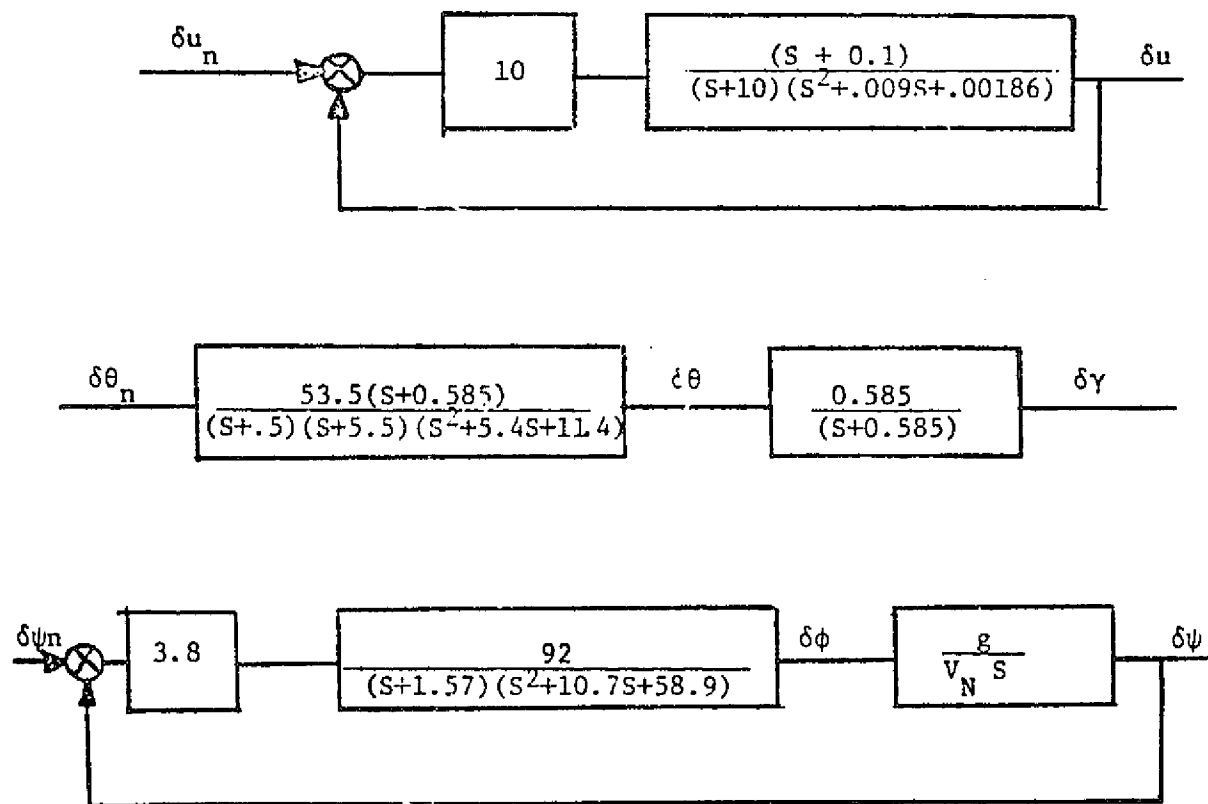


FIGURE 2 BLOCK DIAGRAM OF THE THREE CONTROL LOOPS

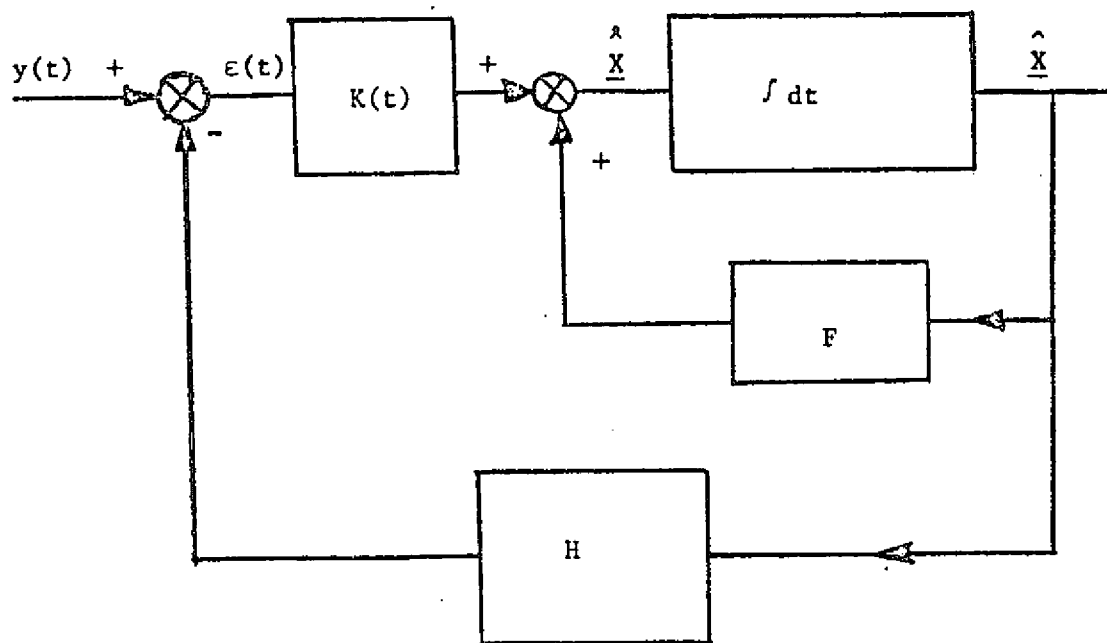


FIGURE 3 LINEAR ESTIMATOR (KALMAN FILTER)

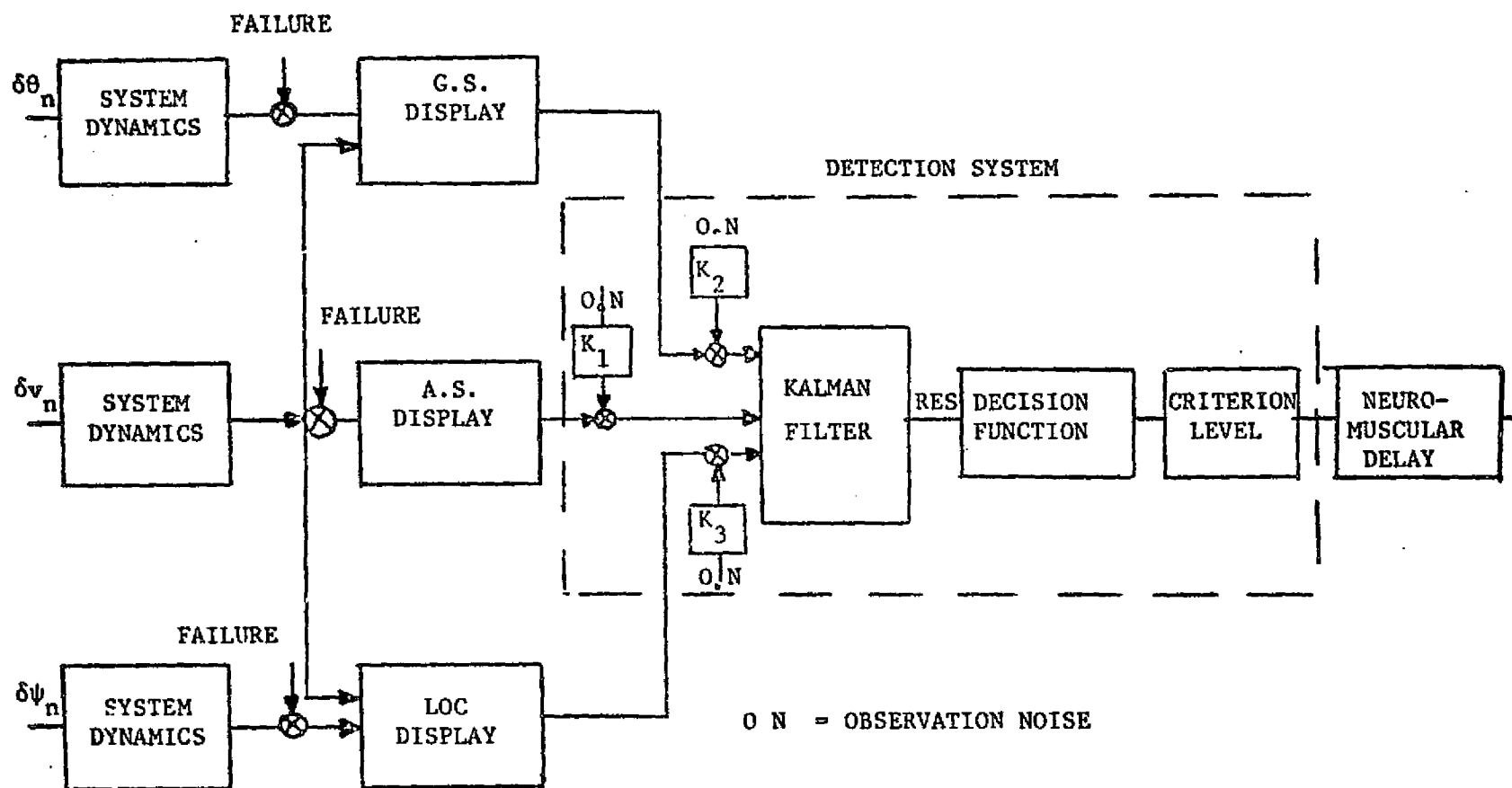


FIGURE 4. FUNCTIONAL DIAGRAM OF FAILURE DETECTION DURING AUTOMATIC LANDING

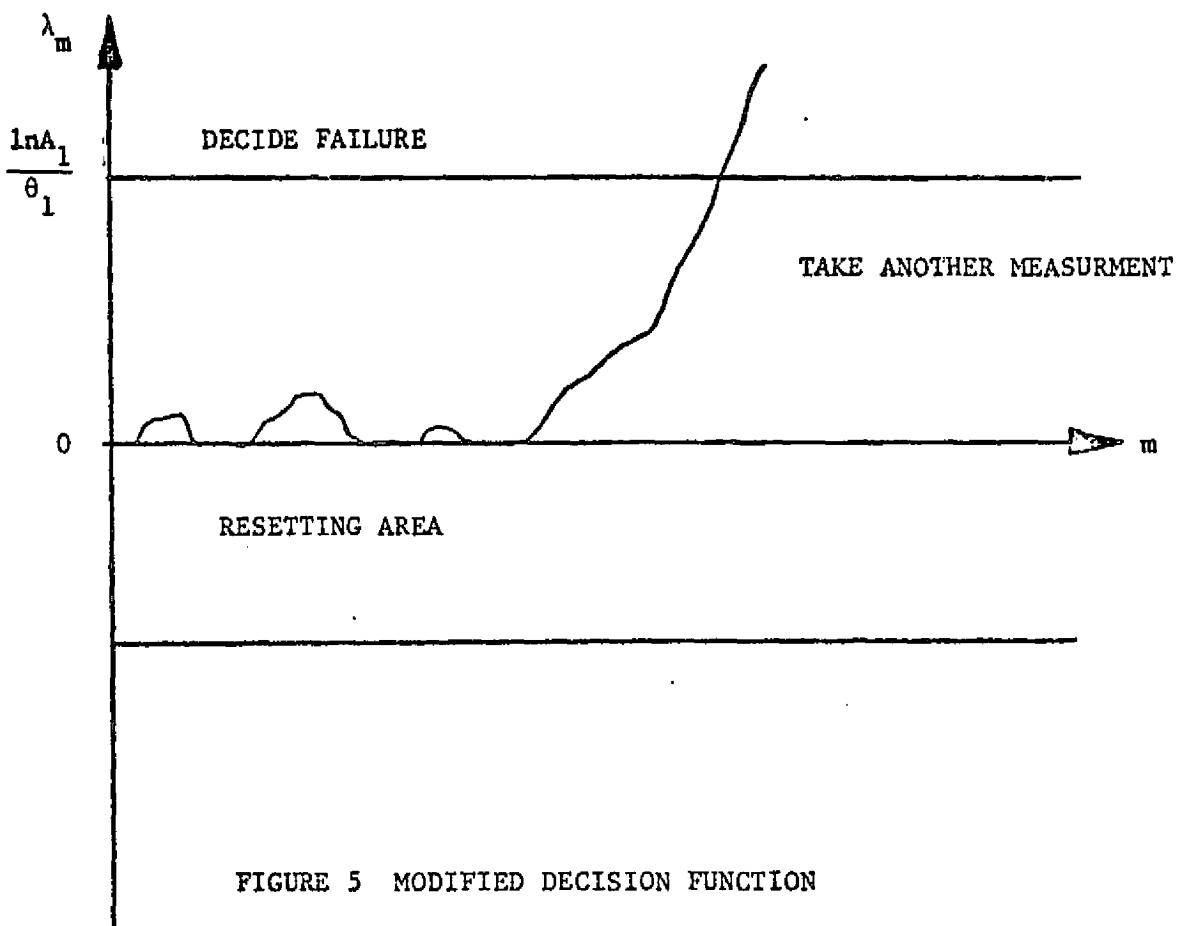


FIGURE 5 MODIFIED DECISION FUNCTION

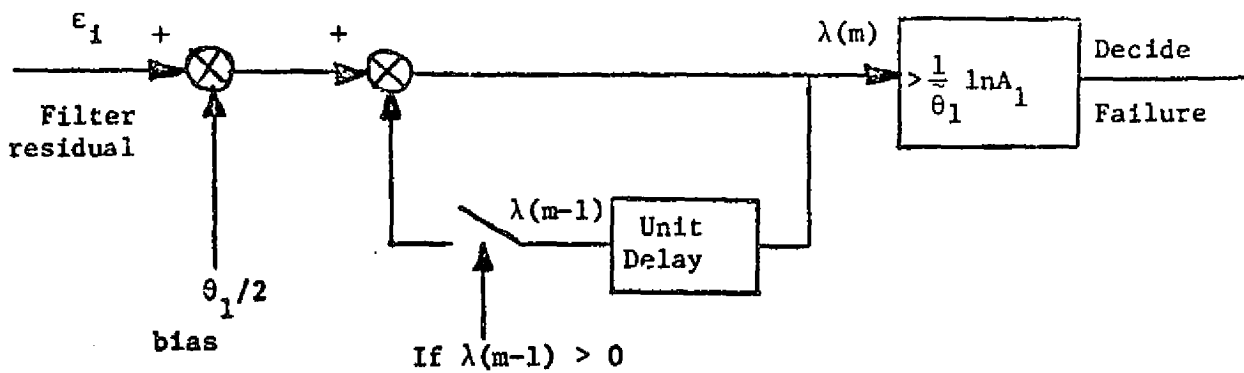


FIGURE 6 DECISION MECHANISM

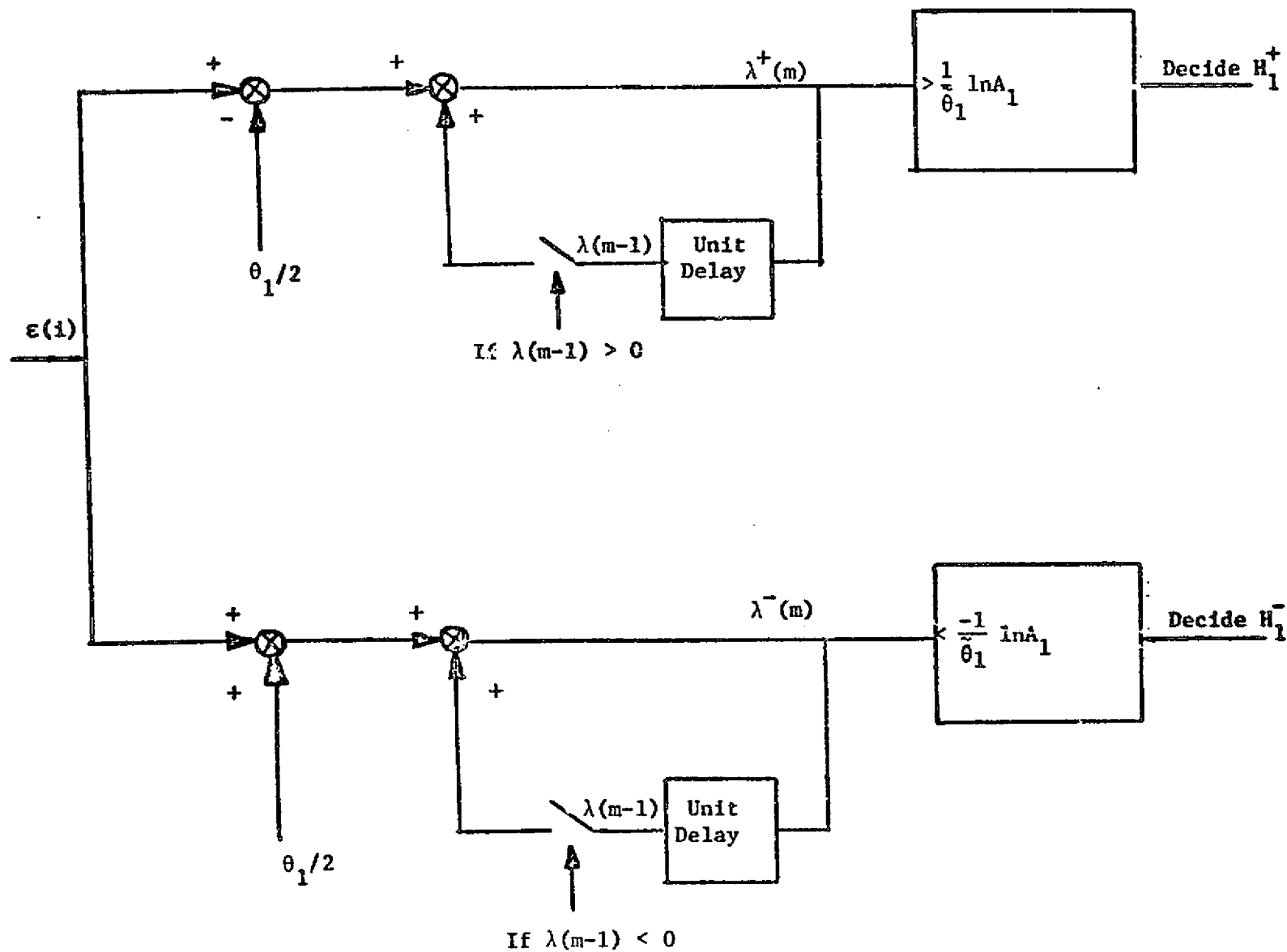


FIGURE 7. COMPLETE BLOCK DIAGRAM OF THE DECISION MECHANISM

FIGURE 8 DETECTION TIMES FOR GS FAILURES (FIRST SET, SUBJECT B.M)

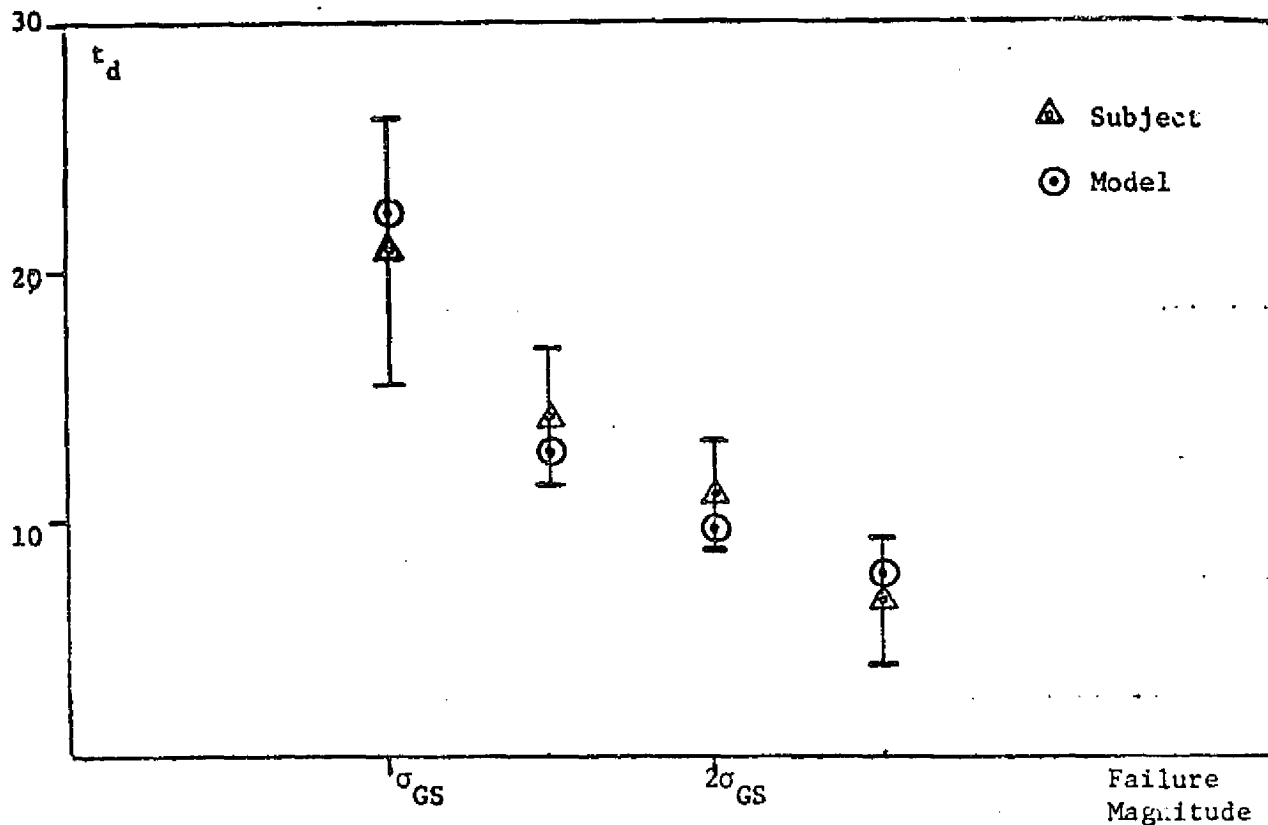
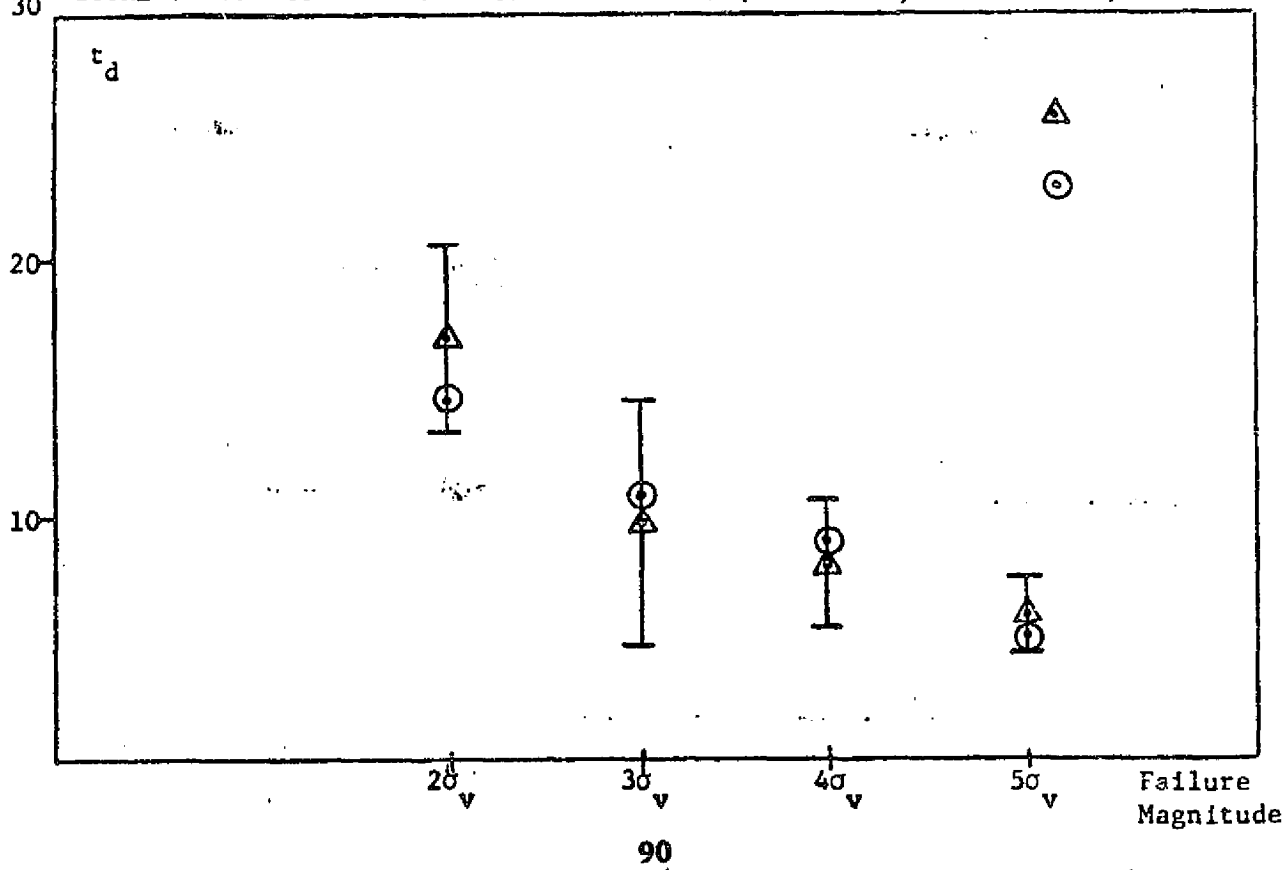


FIGURE 9 DETECTION TIMES FOR AS FAILURES (FIRST SET, SUBJECT B.M)



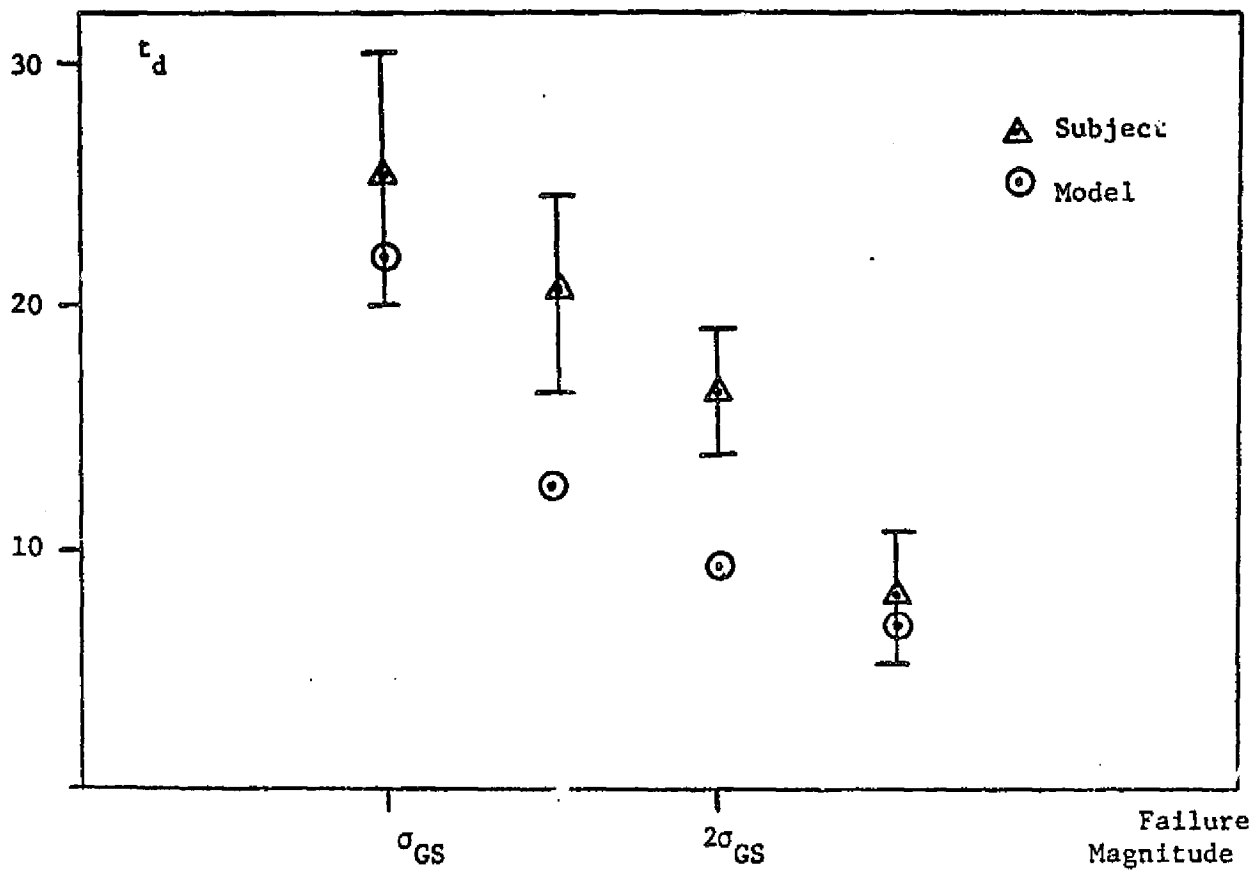


FIGURE 10 DETECTION TIMES FOR GS FAILURES (FIRST SET, SUBJECT C.C)

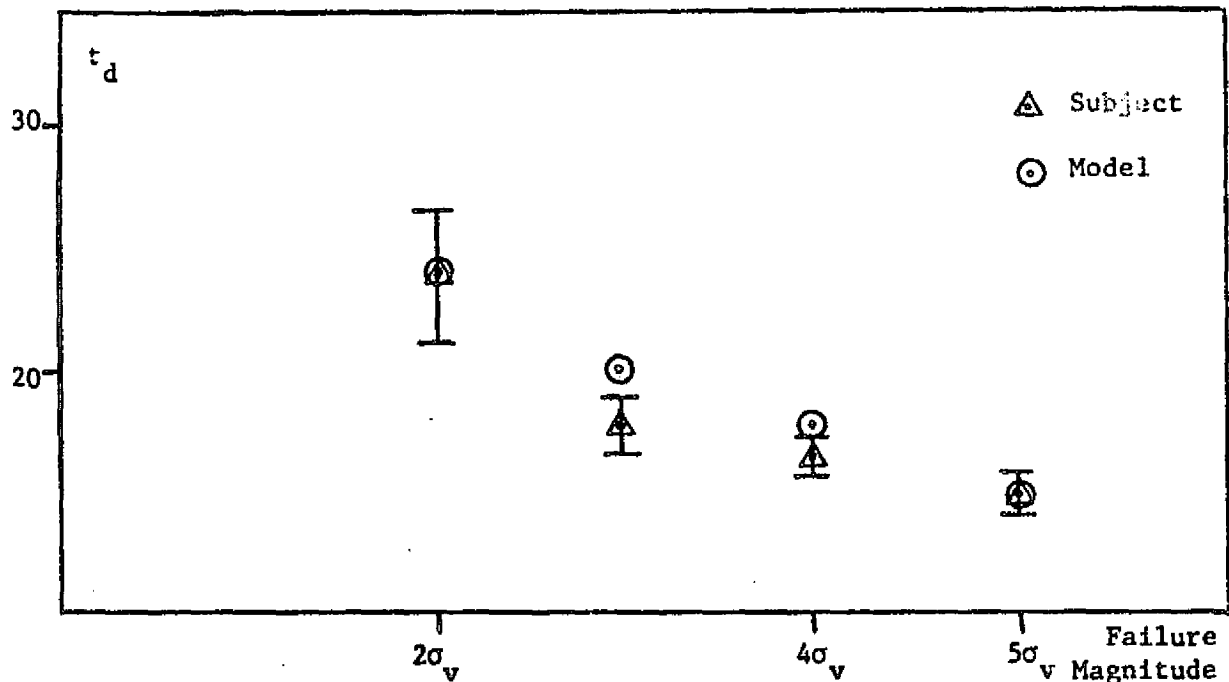


FIGURE 11 DETECTION TIMES FOR AS FAILURES (FIRST SET, SUBJECT C.C)

SUBJECT	INSTRUMENT	FAILURE MAGNITUDE	DETECTION TIMES (SECONDS)			
			C_1	C_2	C_3	C_4
B.M	AS	E(td)	20.8	13.8	10.8	6.3
		σ_{td}	5.9	2.7	2.1	2.7
	GS	E(td)	16.4	9.8	7.7	5.98
		σ_{td}	3.6	4.9	2.4	1.1
C.C	AS	E(td)	25.4	20.8	16.9	8.2
		σ_{td}	5.9	4.0	2.5	2.8
	GS	E(td)	14.0	6.9	6.3	5.0
		σ_{td}	2.8	1.0	0.9	0.9

TABLE I SUBJECTS PERFORMANCE IN EXPERIMENTAL SESSIONS

REFERENCES

1. McRuer, D.T., Jex, H.R., "A review of quasilinear pilot models", IEEE Transactions on HFE 8, 1967.
2. Kleinman, D.L., Baron, S., "Manned vehicle systems analysis by means of modern control theory", Bolt Beranek and Newman Report #1967, 1970.
3. Sheridan, T.B., Ferrel, W.R. Man Machine Systems MIT Press, 1974.
4. Phatak, A.V., Kleinman, D.L. "Current status of models for the human operator as a controller and decision maker in manned aerospace systems" Proceedings of the AGARD Conference, #114, October, 1972.
5. Wald, A. Sequential Analysis, J. Wiley, 1947.
6. Gai, E.G. "Psychophysical models for signal detection with time varying uncertainty", Ph.D. Thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, January 1975.
7. Ephrath, A.R. "Pilot performance in zero visibility precision approach" Ph.D. Thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, June, 1975.
8. Blacklock, J.H. Automatic Control of Aircraft and Missiles, J. Wiley, 1965.
9. Schweppe, F.C. Uncertain Dynamic Systems, Prentice Hall, 1973.
10. Kailath, T. "The innovation approach to detection and estimation theory" Proceedings of the IEEE, 58, 1970.
11. Levison, W.H. "A control theory model for human decision making" Proceedings of the Seventh Annual Conference on Manual Control, 1971.
12. Chien, T.T. "An adaptive technique for a redundant sensor navigation system", C.S. Draper Laboratory Report T-560, 1972.

APPENDIX D

PRECEDING PAGE BLANK NOT FILMED

EXPERIMENTS IN PILOT DECISION-MAKING DURING SIMULATED LOW VISIBILITY APPROACHES

Renwick E. Curry
Man Vehicle Laboratory
Department of Aeronautics & Astronautics
Massachusetts Institute of Technology
Cambridge, MA 02139

John K. Lauber & Charles E. Billings
Man-Machine Integration Branch
Ames Research Center
NASA
Moffett Field, CA 94035

INTRODUCTION

Despite a vast accumulation of operational experience with the conduct of low visibility instrument approaches, little is understood about the decision-making behavior of pilots who fly these approaches. Likewise, there is little information regarding the man, system, and task-related factors which influence this decision-making behavior. Such information is essential for the rational design of new systems, or for the redesign of existing systems in order to correct known deficiencies.

A major problem which has inhibited the study of pilot decision-making behavior in the laboratory has been the unavailability of tasks which incorporate the essential cognitive features of the real task, and which include those motivational or stress-inducing features known to influence decision-making performance. This paper describes a task which was designed to simulate the cognitive features of low visibility instrument approaches, and to produce controlled amounts of subjective stress in pilots serving as subjects in experiments using the task.

Pilot behavior during low visibility instrument approaches can be analyzed into at least two major categories: one is the continuous closed-loop manual tracking behavior necessary to control the aircraft, and the other is the cognitive, decision-making behavior required to make the decision to continue the approach and landing, or to execute a missed approach. It is the second category of behavior with which we are concerned here.

The difficulty of a decision-making task is, in part, determined by the uncertainty of the data used to make a decision. For example, the decision to "go around" is a relatively easy one if, at the missed approach point, there is nothing to be seen outside the aircraft, or if the approach lights and runway have been clearly visible for the last two miles of the approach. It is when the approach lights or runway are barely visible, and then only intermittently, that the decision-making task becomes more difficult.

A second class of variables which are known to influence the outcome of decision-making tasks is best illustrated by the various kinds of psychological stressors acting upon the pilot. Of particular interest here are the pressures perceived by the pilot to complete the approach, to make an on-time arrival, to save fuel, and even to save "face".

We have assumed that it is necessary to use a simulation task which incorporates both kinds of variables, informational and psychological, to successfully study pilot decision-making behavior in the laboratory. The task below was designed to meet those requirements. This paper describes our preliminary experiments in the measurement of decisions and the inducement of stress in simulated low visibility approaches.

METHOD

A schematic of the apparatus as seen by the pilot-subject is shown in Fig.

1. The buttons available to the subject are RVR (to request an RVR reading), turn rate buttons (left, 0, or right), and GA, the go around button to initiate a missed approach.

In the central portion of the CRT is a plan view of the approach. In the lower part of the screen are three dots corresponding to the position of the approaching aircraft (present position, position one second ago, and position 5 seconds ago). In the center of the screen are two pairs of dots corresponding to the middle marker location, equivalent to the 200 foot decision height for a Category I approach. Farther up the screen are the runway outline, threshold, and three pairs of approach lights or lead-in lights. Above that are scores posted for the results of any one trial: on this approach the subject would receive 100 points for a safe landing and -40 points for a missed approach. On the left of the screen is an RVR scale with two indices corresponding to 0 RVR and that for the legal minimum (2400 feet). On the right side of the screen is an altimeter which has a dynamic range of 0 to 220 feet. The pointer indicating altitude is pegged at the upper right until the aircraft nears the middle marker; as the aircraft passes through the middle marker, the indicated altitude passes through 200 feet.

A random wind disturbance from the side (correlation time of 50 seconds) is introduced to provide a moderately easy control task for the pilot. Control is maintained by pushing one of the three turn-rate buttons. The aircraft has the capability of being in either the 0 turn rate (constant heading) or a standard turn rate to the right or left. The pilot's task in these approaches is to "fly" the aircraft through the gate, over the approach lights, and on to the runway. (The aircraft's position shown in Figure 1 is close to the initial condition point.) Only lateral position is important, for if the pilot crosses the extended threshold line, but is not over the runway a crash is recorded. If at any time before the aircraft crosses the extended threshold line, the pilot hits the go-around button, a standard rate left turn is initiated until the heading reaches 60° from "North" at which time the computer program assumes that a missed approach was made.

The runway and approach lights may appear either to the right or left of the middle marker center line, and may be closer or farther away than the nominal position to represent electronic guidance errors. This is the appropriate aircraft-centered view, and simulates the case when one is flying the ILS with needles exactly centered, but finds the runway to the left or right when break-out occurs, and the case when one is either high or low of the indicated altitude.

The slant range "visibility" is included in the program, even though the intensity in the CRT has only two values (on, off). There are 5 "characters" drawn by the PDP-12 graphics system under visibility control: the three pairs of lead-in lights, and the right and left halves of the runway/threshold lights. Should the center of any of these five characters be within a square (centered at the aircraft position) whose half-width is the slant range visibility, then this character will be turned "on" and will be visible. The approach lights are turned off as one gets close to each pair, to simulate their passing

underneath the nose of the airplane; this also prevents the subject from obtaining unrealistic lateral guidance information.

A computer program was written to generate files of approach trajectories and currently has a catalogue of nine approach trajectories. Five of these trajectories have constant (but different) slant range visibilities leading to the following effect: when the middle marker is passed, nothing is in view; soon the first approach light appears, followed by the second and then the third; as the first approach light is neared, it disappears (passes underneath), and then the runway/threshold lights suddenly appear and a safe landing can be accomplished. The decreasing slant range visibility in this group of five trajectories is such that one must proceed farther and farther beyond the middle marker (or below decision height) before the first approach light is sighted. The fifth of these five trajectories is zero-zero visibility, so the approach lights and runway/threshold lights never appear. The other four trajectories correspond to:

- (1) a high visibility approach (runway and approach lights are visible as shown in Figure 1 at all times)
- (2) an extremely optimistic RVR reading, but very low slant range visibility
- (3) passing through a fog bank after initial acquisition of the approach lights: the approach lights and runway lights "drop out", only to reappear after three to four seconds
- (4) fog bank as in (3), but the approach and runway lights do not reappear.

PROCEDURE

In the first set of experiments, we had the following objectives:

- (1) to structure the experimental setting to make the pilot as averse to a crash in the simulator as he would be in real life
- (2) to alter the decision strategies by manipulating the relative values of a landing and a missed approach.

The first objective was desirable to make the decisions as meaningful as possible. After "sacrificing" several pilots, we finally arrived at the following procedure.

As the subject is led into the experimental chamber, he is shown a poster-sized list on the wall of people who have previously been subjects in the experiment. Each subject is listed by name, organization, and score (the total number of points accumulated over the 50 data trials). The first subject on the list was a fictitious one (in this case) and in place of his point score was the word CRASHED in bright red letters. The experimenter writes in the subject's name and organization (e.g. Joe Jones, TWA) and leaves the score column blank. The subject is told at that time that should he crash during the data trials, even if on the first data trial, his services are no longer required. That is, in terms of the experiment, he is "dead".

It was obvious to the subject at this point that he was committed to follow through the experiment, and the idea that he might crash and have the event recorded for all to see had a very noticeable effect on almost all the subjects.

Each of the pilots was allowed a total of 25 practice approaches, the first 10 of which were high visibility approaches so that he could become familiar with the dynamics of the simulation, the wind and turbulence levels, and the layout of the approach lights, runway, etc. After a brief rest period, the pilots participated in the 50 data trials -- it was during these trials that a crash would mean immediate dismissal.

The 50 data trials were composed of six "wild card" approaches, e.g. the approach lights dropping out and then reappearing, or approach lights dropping out and not coming back. The remaining 44 trials were the ones examined for pilot decision-making behavior and consisted of eleven replications of four meteorological visibility levels of 0, 20, 30 and 40 display units, where a visibility of 50 corresponded to having the first approach light come into view as decision height was reached.

These 44 approaches were assigned go-around scores of 100 points for the highest visibility down through -80 points for the lowest visibility and were not assigned randomly, but in a manner which we thought would make the decision most difficult. In general, a negative score corresponded to a low visibility approach and a high positive score to the high visibility approaches.

The data recorded during each approach consisted of a "frame" composed of the current x,y position, the displayed RVR, the state of the turn-rate control and the state of the go-around button. These frames were recorded whenever a control action was executed, an RVR request made, and when the go-around button was pressed. From these data, we can infer the number of times the RVR was requested, the control activity, and the altitude and cross track error at the time of go-around should one be requested.

QUESTIONNAIRE

Thirteen pilot-subjects participated in the test and completed a questionnaire, but as the simulation was changed after the first three pilots, they were not included in the data regarding the simulation itself. Of the remaining 10 subjects, 8 are airline pilots and 2 are IFR rated NASA employees.

The questionnaire consisted of 3 major parts: recent experience in low visibility approaches and missed approaches; fidelity of the decision simulation; and stress ratings for actual low visibility approaches and the simulation.

Recent Experience

Of the 10 pilots completing the questionnaire, 7 had made a total of 37 Category I approaches within the last 12 months (six of these 37 approaches were military approaches). Only 2 missed approaches were made by these 7 pilots. When asked what were the most common causes for executing a missed approach (based on their experience), the 3 most frequently mentioned items were

runway alignment/crosswinds	7 times
visibility	5 times
other traffic	3 times

Simulation Fidelity

The subjects were asked to comment via the questionnaire about the simulator fidelity only with respect to the decision of whether or not to continue an approach. This was done both on a semantic differential scale (Totally Unlike - Completely Identical) and by soliciting comments on the similarities and dissimilarities of the simulation to an actual low visibility approach. The ratings of the subjects are shown in Table I where it is seen that the mean fidelity rating is 5.2 with a standard deviation of 1.87, indicating the usual dispersion in intersubject ratings.

Comments on the similarities of the simulation to a low visibility approach detailed the assimilation of information through different sources (RVR, altitude, and runway alignment). When commenting on the dissimilarities, 3 pilots mentioned the lack of danger ("one will not die if you miss", "... lacks the element of danger"). Two of the pilots mentioned that in a real approach more reliance would be placed on decision height, i.e., that it is a cut and dried decision (a go, no-go situation). Another commented that he felt the reward structure was not correct because in actual flight the rewards for going below minima may be the loss of job, etc, whereas reward here is a higher point count.

There were other comments made about dissimilarities of the simulator and the actual approach: three pilots mentioned that the visual cues were different, and one pilot mentioned the fixed turn rate characteristics of the simulator. Those were offered even though the question specifically asked about the similarities of decision making; either the questions were misunderstood or these factors really do influence the decision. In either case, we felt that these latter two factors are of secondary importance in the light of the other dissimilarities mentioned by the pilots.

Stress Ratings

The pilots were asked to rate the stress of the experimental task and an actual low visibility approach on a semantic differential scale (Not at all stressful - Extremely stressful); the results are shown in the other columns of Table I. We have added columns showing the difference in stress rating, and the simulator stress (rating) as a fraction of the actual stress (rating). Of these 10 subjects, three felt that the simulator was at least as stressful as an actual low visibility approach. At the other extreme, is subject number six who reported that the simulator "lacks the element of danger".

RESULTS

One pilot misunderstood the instructions because he initiated a missed approach after safely crossing the threshold many times during the 50 data trials and therefore received less than full point score. His data were not analyzed.

Learning

A statistical test was used to ascertain whether or not a learning effect was present for the subject group by performing an analysis of variance on the

pull-up altitude for those 11 approaches made under zero/zero visibility conditions. This analysis of variance included the approach number as a covariate and, if significant, would suggest a linear trend in pull-up altitude with trial number. The results of this analysis of variance indicated that the covariate of approach (trial) number did not contribute a statistically significant linear component to the pull-up altitude. Thus the linear trend was ignored in the remainder of the analyses.

Classification of Approaches into Land/Go-Around

The 44 approaches for each of the 9 pilots were examined for classification into the classes of land/go-around using a stepwise discriminant analysis program (BMD 07M). The variables included for the selection in the stepwise discrimination were the following:

$(V/V_{MAX})^n$ $n = 1, 2, 3$	Actual visibility on the approach ($0 \leq V/V_{MAX} \leq 1$)
$(S/S_{MAX})^n$ $n = 1, 2, 3$	Score increment for selecting a go-around ($-0.80 \leq S/S_{MAX} \leq 1.0$)
$(V/V_{MAX})(S/S_{MAX})$	Interaction between visibility and score
LOC	Visible localizer deviation at go-around

The linear, quadratic, and cubic component of visibility and go-around score are self explanatory and the interaction term was included to test for its significance. The localizer deviation was also included and taken to be the maximum visible localizer deviation. It was set to 0 on the 0/0 approaches since it would not be available to the pilot, and was also set to zero on the approaches which were successfully completed.

The significant variables selected by the stepwise discriminant program are shown in Table II for each subject; these coefficients have been normalized so that the coefficient of V/V_{MAX} is unity. In this table, an increase in the discriminant function will put the approach into the "land" class. The resulting classification using this discriminant function is also shown in the table and it gives very good results on these data (although one must be cognizant that the classification is performed on the data from which the discriminant function was determined.)

The major points to be ascertained from the table are first, the go-around score was almost useless as a basis for discriminating among approaches with the exception of Subject number 6. (This particular discriminant function must be treated with care since it incorporates almost all the variables and includes a sign of V/V_{MAX} which is opposite from all the other discriminant functions.) The second point of interest is the nearly equal coefficient on the quadratic component of visibility, indicating the decrease in effectiveness of actual visibility in classifying an approach into "land". Thus the

contribution of visibility to the discriminant function varies from 0 (for V/V_{MAX} of 0) to 0.4 (for $V/V_{MAX} = 1.0$) for most subjects.

The coefficient for the localizer deviation can be used to determine the sensitivity of the cross-track error in classifying approaches into "land" or "go-around". This result is shown in the right hand column of Table II and is the localizer error (in degrees) which has the same effect on the discriminant function as a full-range change in slant range visibility. This gives the importance of localizer error relative to visibility in determining the classification.

A DYNAMIC DECISION MODEL

In this section, we briefly describe a decision-theoretic approach to the modelling of pilot decisions during the simulation of low visibility approaches. A straightforward application of the Subjective Expected Utility (SEU) theory is complicated by the dynamic character of the decisions since the theory relates to static decision alternatives, rather than the everchanging situations experienced by pilots. Nonetheless, we have developed an extension (based on SEU models) which appears to be plausible, and it is one which we think will be a valuable tool in further investigations.

The dynamic decision model is based on the assumption of the existence of a decision function which can be written

$$D(V, S, L, h) = \bar{U}_{LAND} - \bar{U}_{GA} \quad (1)$$

where D is the decision function, an explicit function of visibility (V), go-around score/incentive (S), the cross-track or localizer deviation (L), and the current altitude (h). This decision function is the comparison of the SEU for \bar{U}_{LAND} and the SEU for making a missed approach or a go-around \bar{U}_{GA} . Under the assumption that the utilities for landing, crashing, and going around are independent of the probabilities, Subjective Expected Utilities take the form

$$\bar{U}_{LAND} = P_{LAND}(V, L, h)U(LAND) + P_{CRASH}(V, L, h)U(CRASH) \quad (2)$$

$$\bar{U}_{GA} = 1 \cdot U(GO AROUND) = U(S) \quad (3)$$

where P_{LAND} and P_{CRASH} are the subjective probabilities for landing and crashing and $U(\cdot)$ is the corresponding utility. These expressions show the dependence on the approach variables V , L , h and the incentive for going around S .

A schematic plot of the decision function and how it might change with altitude is shown in Figure 2. We have displayed possible variations of \bar{U}_{LAND} during an approach and its comparison with \bar{U}_{GA} which remains constant throughout the approach. If at any time the SEU of landing becomes less than that of going around, the decision is made to initiate a missed approach.

A missed approach, denoted by the solid line of Figure 2, is a sketch of how the SEU of landing might behave during an approach for which the approach lights are never sighted. The SEU of landing decreases with altitude because the subjective probability of landing is decreasing (and that of crashing is increasing) until, at point A, a missed approach is initiated. The SEU of an

approach shown by the dotted line starts out higher than the previous approach (perhaps because of a larger reported RVR) and it, too, decreases until point B when the approach lights are sighted. This causes an immediate jump in the probability of landing (hence the step change in SEU of landing) and from this point gradually increases until point C when the landing is successfully accomplished. If, on the other hand, the aircraft starts deviating from the localizer (say) at point D, and the pilot has difficulty in stabilizing the approach, then the subjective probability for landing will decrease causing a decrease in the SEU for landing until point E where a missed approach is initiated because of misalignment with the runway.

The point of greatest interest, of course, is at the instant of deciding to initiate a missed approach. At this instant the SEU of landing and go-around are equal, i.e.

$$U_{\text{LAND}}(V, L, h^*) = U_{\text{GA}}(S) \quad (4)$$

where h^* is the altitude at which the go-around decision was made and is written

$$h^* = g(V, L, S) \quad (5)$$

To test for the possible existence of such a relation, we performed a stepwise regression (using BMD 02R) on the go-around score (S/S_{MAX}), score/visibility interaction ($S/S_{\text{MAX}})(V/V_{\text{MAX}}$), and the localizer deviation (L). The main effects of meteorological visibility were not included because the majority of the go-around decisions were made under 0/0 visibility conditions. Table III shows the regression coefficients selected by the stepwise regression program; those coefficients which have a non-zero value as indicated by Student's t test are indicated with an asterisk. Note that the data for subjects 2,3,4 are not included because the stepwise regression program did not find a significant regression on the variables indicated. The multiple regression coefficient is highest for those cases for which few go-around decisions were made during approaches when any visibility existed (recall that 11 of the approaches shown were made under 0/0 conditions). While the coefficients indicate the type of behavior one would expect, e.g. an increased decision altitude due to increased go-around scores, these data must be considered preliminary because of the experimental design (see the discussion section).

These results are quite encouraging and indicate that the subjectively expected utility model proposed here may lead to a valuable viewpoint from which to examine pilot decision making during low visibility approach.

SUMMARY

Stress:

One of the major goals outlined for this preliminary set of experiments was to investigate methods of applying psychological stress analogous to the stress of an actual low visibility approach. It was found that the stress rating in the simulation, as reported on semantic differential scales, was an average of 0.8 times the stress rating of an actual low visibility approach. The success in applying the stress was not uniform, however, for several

subjects reported they "would not die" if they crashed in the simulation. Other subjects remarked that descending below the decision height of 200 feet would result in censure by regulatory or company authorities; this seemed as important to them as the prospect of an (unlikely) crash. Thus it may be meaningful to include another penalty in the simulation; if the subjects are "caught" descending below minimums, they will be penalized (say) by a score equivalent to two or three landings.

Go-Around Incentives

The range of go-around scores did not induce as much behavioral change as was expected, and the results of the experiment indicate that considerably more score differential will be required to induce pilots to initiate a missed approach. For example, when offered 100 points for the no-risk go-around or 100 points for a successful landing, most pilots initiated the approach and almost all continued until touchdown, even though a riskless go-around was available. This suggests that the utility of landing is strongly affected by the accomplishment of this feat, or that the level of risk taking in a go-around is too low and this alternative does not present enough of a challenge to the pilots.

Behavioral Models

One model of behavior which may apply here is the Theory of Achievement Motivation (Atkinson, 1964). This model is of a form similar to the SEU model described above except that the utilities depend on the subjective probabilities: for example, the utility of succeeding at an easy task is low, while succeeding on a difficult (high risk) task is high. Conversely, the utility of failing on an easy task is low (one loses face), whereas there is no disutility (loss of face) on failing to succeed on a risky task. Risk taking behavior is said to be determined by two personality traits; need for achievement and test anxiety. The Theory of Achievement Motivation predicts that those individuals with a high need for achievement and a low test anxiety will take an intermediate level of risk, whereas individuals with a low need for achievement and high test anxiety will take extreme levels of risk: a low level of risk to insure success, or a high level of risk in which success is not really expected. Atkinson draws analogies to aspirations in employment as well as more quantitative behavioral tests such as the "ring toss" experiment. We have conducted some informal experiments at MIT using a ball-toss paradigm involving two levels of difficulty; the results of this undergraduate student project will be reported elsewhere.

A preliminary attempt was made to apply the Theory of Achievement Motivation to the experimental results described above. An obvious measure of test anxiety is the stress ratio recorded by the subjects (stress in the simulation/stress in actual low visibility approach). Measures of success and need for achievement are ambiguous, however, since point score and number of landings may be considered as measures of both.

Nonetheless, if one considers (a) the final score as a measure of success; (b) the stress fraction as the measure of test anxiety; and (c) the number of landings as the measure of the need for achievement (e.g. sticking with an

approach through approach light dropout, etc.), then classification of the pilots on this basis (b) and (c), indicates the following: the pilots with the lower success level (lower score) exhibited a low need for achievement and high test anxiety (as the theory predicts), but the pilots with a higher success level (higher score) exhibited not only a high need for achievement but a higher stress level (rather than the theoretically predicted low stress). Although there are not enough subjects to validate this conjecture statistically, it may well be that the Theory of Achievement Motivation is not applicable in those cases where the result of the failure is catastrophic, and that modification to the theory may be required for situations such as are considered here.

In summary, both a dynamic version of Subjective Expected Utility Models and (a modified) Theory of Achievement Motivation may be useful in describing decision behavior of pilots in a simulated low visibility approach.

REFERENCES

Atkinson, J.W. An Introduction to Motivation, Princeton, N.J.:Van Nostrand, 1964.

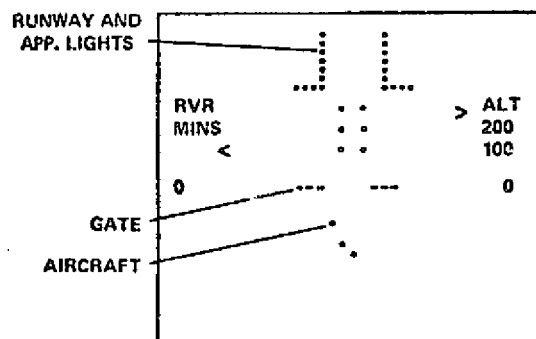


Figure 1 Subject's display panel

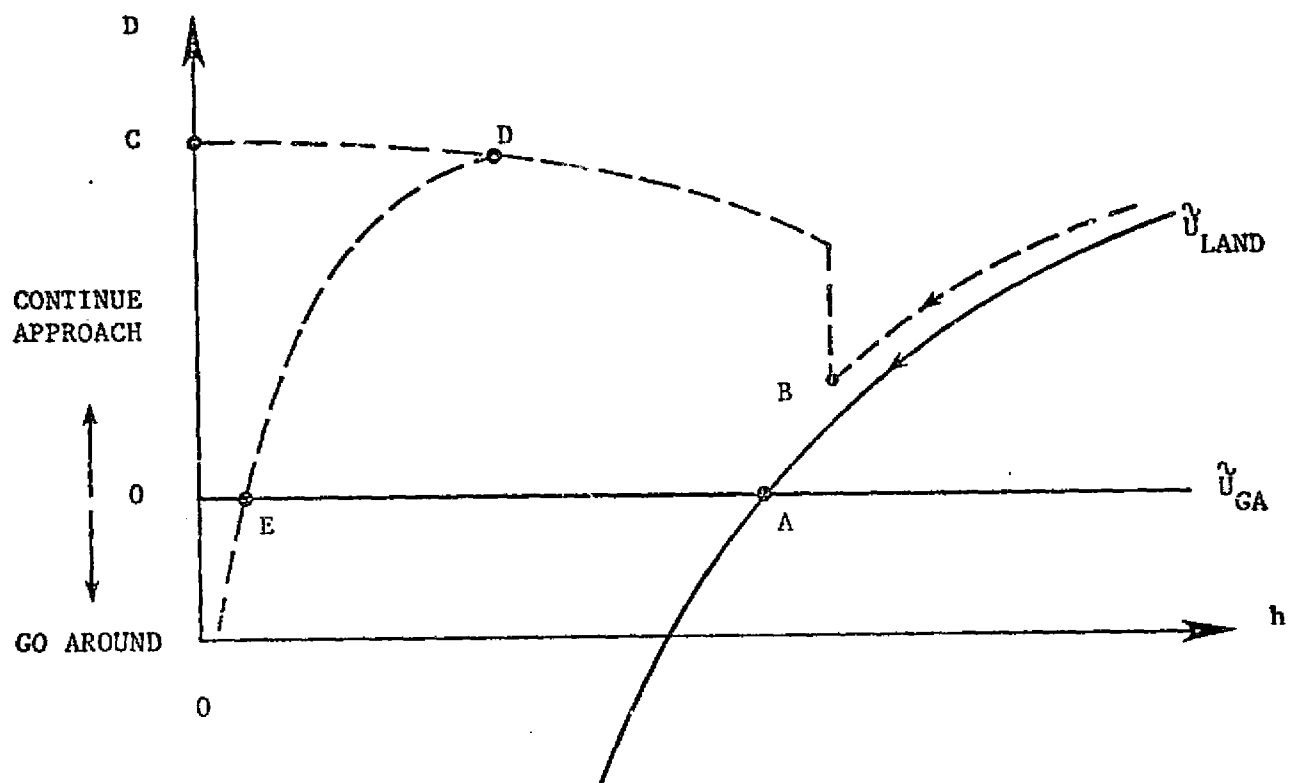


FIGURE 2. Decision function versus altitude

Subject/ Organiza- tion	Simulator Fidelity Rating	Stress Rating -Actual Approach	Stress Rating Simulator	$S_{SIM} - S_{ACT}$	$\frac{S_{SIM}}{S_{ACT}}$
1/A	7	3	2	-1	.67
2/B	7	4	6	2	1.50
3/C	3	7	3	-4	.43
4/B	3	8	5	-3	.62
5/B	7	8	6*	-2	.75
6/A	4	7	2	-5	.28
7/A	7	6	6	0	1.00
8/C	6.5	5.5	5.5	0	1.00
9/D	2.8	7	4	-3	.57
10/D	5	8	6.5	-1.5	.81
Mean	5.23	6.35	4.60	-1.75	.763
S.D.	1.87	1.73	1.73	2.01	.344

* Indicated a change to 2 later in the trials

TABLE I Semantic differential ratings of simulator fidelity and stress

SUB	$\left(\frac{V}{V_{MAX}}\right)$	$\left(\frac{V}{V_{MAX}}\right)^2$	$\left(\frac{S}{S_{MAX}}\right)$	$\left(\frac{S}{S_{MAX}}\right)^2$	$\left(\frac{S}{S_{MAX}}\right)^3$	$\left(\frac{S}{S_{MAX}}\right)\left(\frac{V}{V_{MAX}}\right)$	LOC	CLASSIFICATION		LOCALIZER DEVIATION EQUIVALENT TO FULL VISIBILITY CHANGE
								GROUP	L G	
1	1.0	-0.155				-0.352	-0.0185	23 1	0 20	45.6°
2	1.0	-0.650					-0.00432	31 0	0 13	81.0°
3	1.0	-0.507	-0.065			-0.180	-0.0109	33 0	0 11	45.0°
4	1.0	-0.673					-0.0045	29 1	0 14	72.6°
5	1.0	-0.746					-0.00695	30 1	0 13	36.1°
6	-1.0	1.34	0.650	0.14	-0.417	-0.793	-0.0066	23 0	0 21	51.2°
7	1.0	-0.619					-0.00848	29 0	0 15	44.3°
8	1.0	-0.695	-0.101				-0.00679	27 1	0 16	44.9°
9	1.0	-0.635					-0.0088	30 1	0 13	41.5°

Table II Coefficients of Discriminant Function (Normalized such that $|V/V_{MAX}| = 1.0$)

SUB	CONST	$\frac{S}{S_{MAX}}$	$\left(\frac{S}{S_{MAX}}\right)^2$	$\left(\frac{S}{S_{MAX}}\right)^3$	$\left(\frac{S}{S_{MAX}} \frac{V}{V_{MAX}}\right)$
1	88.6	124*	45*	-106*	90.1
5	17.3	-143.3*	-136.7*		318*
6	44.9	30.6	66.5*	19.7	-62.8
7	101.0	245*	41.3*	-185*	
8	71.3	111.6*	10.1	-93.6*	-140*
9	-17.4	-78.5		79.1*	-44.4*

	LOCALIZER	MULTIPLE ρ	σ_e (ft)	N
1	1.41*	0.77	16.1	21
5	-0.070	0.92	10.5	14
6		0.69	26.5	21
7	-3.83*	0.97	5.0	15
8	-0.32	0.91	10.5	17
9	1.02	0.94	6.2	14

TABLE III Regression coefficients for altitude of go-around decision

APPENDIX E

PRECEDING PAGE BLANK NOT FILMED

Program Abstracts/Algorithms

A multinomial maximum likelihood program (MUNOML)

RENWICK E. CURRY

Man-Vehicle Laboratory
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

In our research on modeling sensory and decision phenomena, we were soon confronted with the task of evaluating both old and new models using both old and new data. Rather than design an ad-hoc estimation program for each new model, as is typically done, we developed an "executive" program that provides a general method for estimating parameters and simultaneously provides flexibility for accommodating new models with a minimum amount of programming. Our experience with canned computer programs has been equivocal, so we decided to provide only the general framework and let the user accomplish the objective of estimating parameters for his particular model by writing a new subroutine within the constraints of the executive program.

The most common class of distributions for which parameters must be extracted are multinomial distributions resulting from a stimulus-response classification, e.g., binary responses (YES-NO or two alternative forced choice methods), the method of successive categories (rating scales), or transition probabilities in a Markov chain. Although a number of methods exist for estimating such parameters (Restle, 1971) we have chosen the maximum likelihood method and have implemented the scoring of Rao to adjust the parameters from one iteration to the next. We have chosen the maximum likelihood (ML) method because (1) it is a member of the class of consistent asymptotically normal estimators (CAN) under suitable conditions (Rao, 1973); (2) it will easily handle situations in which all the responses fall into one category; (3) there are many situations in which the maximum likelihood estimator can be shown to yield unique estimates for parameters (other estimation techniques may have this property as well); and (4) it is the only one exhibiting first order efficiency (Rao, 1973).

Other authors have successfully used the maximum likelihood method with Rao's scoring method in signal detection type models (e.g., Dorfman and Alf, 1968; Dorfman and Alf, 1969; Dorfman, Beavers, & Saslow, 1973; Grey & Morgan, 1972; Ogilvie & Creelman, 1968). This program differs primarily in its ability to estimate parameters for a wide variety of models.

A BRIEF REVIEW OF MAXIMUM LIKELIHOOD ESTIMATION

Necessary Conditions for a Maximum

In this section, we summarize the pertinent aspects of the maximum likelihood (ML) estimators pertaining to the practical considerations of carrying out the estimation process. [See the excellent descriptions by Rao (1973) and Kendall & Stuart (1967) for a more detailed discussion of ML estimators.] We will formulate the problem as it is usually encountered in behavioral research, that is, in the form of a multinomial distribution containing parameters to be estimated. We specify by P_{ij} the theoretical conditional probability of observed response

Sponsored by NASA Grant NGR 22-009-733, Man-Machine Integration Branch, NASA-Ames Research Center.

$j(j = 1, \dots, n)$ given stimulus $i(i = 1, \dots, m)$. Thus, we have a series of multinomial distributions (multimultinomial), and it is assumed that the responses in each of these distributions are independent. Under these conditions, the likelihood function for the multinomial distribution is proportional to $L(x)$ (Rao, 1973).

$$L(x) = \prod_{i=1}^m \prod_{j=1}^n P_{ij}^{r_{ij}}(x), \quad (1)$$

where x is the 2 -dimensional parameter vector to be estimated and r_{ij} is the number of " j " responses observed when the stimulus is specified at condition " i ."

As is usual, we will consider the natural logarithm of the likelihood function, since it is easier to deal with numerically and analytically, and its maximum occurs at the same point as that of the likelihood function.

$$\ln L(x) = \ell(x) = \sum_{i=1}^m \sum_{j=1}^n r_{ij} \ln P_{ij}(x). \quad (2)$$

The first order necessary conditions for an extremum point of the likelihood function are given by the (components of the) vector equation:

$$\frac{\partial \ell}{\partial x} = \sum_{i=1}^m \sum_{j=1}^n r_{ij} \frac{1}{P_{ij}(x)} \frac{\partial P_{ij}}{\partial x} = 0. \quad (3)$$

Whether any extremum point satisfying (3) is a local maximum, minimum, or saddle point is found by examining the eigenvalues of the matrix of second partial derivatives or the Hessian matrix

$$\begin{aligned} \frac{\partial^2 \ell}{\partial x^2} = & - \sum_{i=1}^m \sum_{j=1}^n \frac{r_{ij}}{P_{ij}^2(x)} \left[\frac{\partial P_{ij}}{\partial x} \right] \left[\frac{\partial P_{ij}}{\partial x} \right]^T \\ & + \sum_{i=1}^m \sum_{j=1}^n \frac{r_{ij}}{P_{ij}(x)} \frac{\partial^2 P_{ij}}{\partial x^2}. \end{aligned} \quad (4)$$

Finding the Maximum

Since the necessary conditions for an extremum (3) are rarely, if ever, analytically tractable, an iterative approach is called for. The general form of the parameter updating technique usually takes the form

$$x_{i+1} = x_i + A_i^{-1} \frac{\partial \ell}{\partial x}, \quad (5)$$

where x_i is the parameter estimate at the i th iteration, $\partial \ell / \partial x$ is the gradient evaluated at the point x_i , and the matrix A_i determines the algorithm by which one seeks the maximum of the likelihood function. One approach, Rao's scoring method, is to construct the matrix A from an approximation to the information matrix

$$A = \sum_{i=1}^m \sum_{j=1}^n \left[\frac{r_{ij}}{P_{ij}^2(x)} \right] \left[\frac{\partial P_{ij}}{\partial x} \right] \left[\frac{\partial P_{ij}}{\partial x} \right]^T. \quad (6)$$

Note that the matrix A_i is the negative of the Hessian matrix (4) with the last set of terms omitted. An advantage of the method of scoring is that the matrix A_i can be calculated from the gradient itself, and the second partial derivatives of P_{ij} are not needed. When inverted, as it must be for the iterations, the inverse becomes the Cramer-Rao lower bound covariance matrix of estimation errors in the parameter x . See Curry (Note 1) for a more detailed discussion of these points.

MUNOML: The Executive Search Program

As can be seen from Equations (3), (5), and (6), the only quantities needed to carry out the search function are the $m \times n$ quantities P_{ij} , and the $m \times n \times 2$ quantities $\partial P_{ij} / \partial x_k$. We have written MUNOML to carry out the search operation using only the numerical values of the probabilities and the partial derivatives. These numerical values are provided by a user supplied subroutine which is called once in each iteration cycle. This provides an extremely flexible tool for estimating parameters of grouped data, i.e., for multinomial distributions. Thus, it will accommodate any behavioral model in which the stimuli and responses are drawn from discrete element sets as long as $\partial P_{ij} / \partial x_k$ is continuous. The theoretical form of the underlying distributions within the model are contained in the subroutine. The range of distributions and models is almost limitless: One may treat continuous sensory models (signal detection theory); Thurstonian scaling; discrete sensory models (rectilinear ROC curves); predictions of judgments and choice by the method of successive categories (more Thurstonian scaling). The distributions may belong to the same location/scale family (Gaussian, logistic, arc sine, etc.) or they may involve imbedded parameters (e.g., the beta distribution).

In addition to the maximization of the likelihood function with respect to the parameters, we have included a χ^2 goodness-of-fit test for the final parameter estimates, given by

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^n \frac{(r_{ij} - r_i P_{ij})^2}{r_i P_{ij}}, \text{ where } r_i = \sum_{j=1}^n r_{ij}; \quad (7)$$

$$df = m(n-1) - \ell$$

The degrees of freedom are obtained by finding the number of independent measurements and subtracting the number of parameters being estimated. We also have included another χ^2 computation in which all cells for which the expected number of responses is less than four are pooled with adjacent cells and the degrees of freedom are reduced accordingly.

MODELS IN SIGNAL DETECTION AND RECOGNITION**Psychological Continuum (Thurstonian) Models**

Rating scales—The method of successive categories: In the method of successive categories (Bock & Jones, 1968), or the rating method in psychophysics (Green & Swets, 1966), the observed measures are P_{ij} , the probability of eliciting response R_j given stimulus S_i . These sensory-continuum models assume that each presentation of a stimulus maps to a point on an internal continuum, and these same probabilities, as predicted by the models, take the form

$$\begin{aligned} P_{ij} &= P(R_j/S_i) = P(c_j \leq u < c_{j+1}/S_i) \\ &= P\left(\frac{c_j - \mu_i}{\sigma_i} \leq \frac{u - \mu_i}{\sigma_i} < \frac{c_{j+1} - \mu_i}{\sigma_i}\right) \\ &= F(Z_{i,j+1}) - F(Z_{i,j}) \\ i &= 1, 2, \dots, m \quad j = 1, 2, \dots, n. \end{aligned} \quad (8)$$

$F(\cdot)$ is the theoretical distribution function; and the unit deviates are

$$Z_{i,j} = \frac{c_j - \mu_i}{\sigma_i} = b_i c_j - a_i. \quad (9)$$

Here c_j is the j th criterion level; σ_i and μ_i are the mean and standard deviation of the discriminial dispersion of the i th

stimulus; and $b_i = 1/\sigma_i$ and $a_i = \mu_i/\sigma_i$ are an alternative set of parameters introduced for convenience.

To remove the ambiguity of the origin and scale factor of the distributions, we assign

$$\begin{aligned} b_i &\equiv 1, \text{ which fixes scale} \\ a_i &\equiv 0, \text{ which fixes origin.} \end{aligned} \quad (10)$$

Equations (10) are particularly useful when using rating scale methods in psychophysics, since one can arbitrarily assign the noise process to stimulus S_1 , and all distributions are scaled with noise standard deviation so that b_i becomes the ratio of noise standard deviation to signal standard deviation. In the equal-variance case, a_i is the normalized separation between "noise" and "signal" distributions, often called d'_i .

Multiple SNR pairs with equal variance: One of the new models that we have investigated is for decision behavior at multiple signal-to-noise ratios (SNR) in which the experimental conditions were such that an equal-variance Gaussian model is apropos. Rather than work with criterion cut-off points, c_j it is more meaningful to deal with the log likelihood ratio which, for equal-variance Gaussian distributions, can be shown to be

$$\lambda(u) = d'_i [b_i u - 1/2(a_i + a_{i+1})] \quad i = 1, 3, \dots, m-1. \quad (11)$$

In the above expression, we have used the convention that the stimuli are ordered by equal SNR pairs, i.e., S_1 and S_2 have the same SNR, S_3 and S_4 are at another SNR, etc. When this expression for the log likelihood ratio is substituted into Equation (9) for the deviates with $u = c_{ij}$, we obtain

$$\begin{aligned} Z_{i+1,j} &= \lambda_{i,j}/d'_i + d'_i/2 \quad i = 1, 3, \dots, m-1 \\ Z_{i,j} &= \lambda_{i,j}/d'_i - d'_i/2 \quad j = 2, 3, \dots, n. \end{aligned} \quad (12)$$

If one assumes a constant likelihood decision rule, then $\lambda_{ij} = \lambda_j$ for all i .

Discrete Sensory Models

The discrete sensory models lead to the prediction of rectilinear ROC curves (e.g., see Krantz, 1969 & Luce, 1963). These may be expressed by

$$P_H = p_0 + \sum_{i=1}^N \Delta p_i S(c-i) \quad P_{FA} = q_0 + \sum_{i=1}^N \Delta q_i S(c-i) \quad (13)$$

with the constraint

$$p_0 + \sum_{i=1}^N \Delta p_i = q_0 + \sum_{i=1}^N \Delta q_i = 1,$$

where N is the number of limbs in the ROC; P_H and P_{FA} are the theoretical values for the probability of a hit and false alarm; p_0 , Δp_i , q_0 , Δq_i describe the form of the ROC curve; c is the criterion level; and $S(\cdot)$ is a saturation function.

$$S(x) = \begin{cases} 0 & x \leq 0 \\ x & 0 < x < 1 \\ 1 & 1 \leq x \end{cases} \quad (14)$$

Equations (13) and (14) are suitable for use with the method of scoring as an iterative technique since the gradient is reasonably

well behaved, although only piecewise continuous with respect to c .

SUMMARY

We have found MUNOML to be extremely useful in day-to-day operation since it provides the maximum amount of flexibility with a minimum amount of duplicated effort. This is especially true when minor changes to a model's structure are made because there will be a great deal of unchanged code within the user supplied subroutine.

MUNOML has excellent convergence qualities and rarely goes beyond 10 iterations; when it does, it is usually because of programming errors in the calculation of the probabilities or their derivatives.

We are collecting a library of subroutines for use with MUNOML. There are currently subroutines for calculating parameters for a successive categories model and for models based on the theory of signal detectability: unequal variance ROC parameters from binary responses; equal-variance signal-to-noise ratio pairs; constant likelihood ratio decision rules; and Neyman-Pearson decision rules (using objective and subjective probabilities). Complete listings and/or card decks (about 500 cards in all) are available on request. Operational details are described in Curry (Note 1).

REFERENCE NOTE

1. Curry, R. E. MUNOML: A multinomial maximum likelihood program for behavioral research. Man-Vehicle

Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1974.

REFERENCES

- Bock, R. D., & Jones, L. V., *The measurement and prediction of judgement and choice*. San Francisco: Holden-Day, 1968.
- Dorfman, D. D., & Alf, E., Jr. Maximum-likelihood estimation of parameters of signal-detection theory—a direct solution. *Psychometrika*, 33, 117-124, 1968.
- Dorfman, D. D., & Alf, E., Jr. Maximum-likelihood estimation of parameters of signal detection theory and determination of confidence intervals—rating method data. *Journal of Mathematical Psychology*, 6, 487-496, 1969.
- Dorfman, D. D., Beavers, L. L., & Saslow, C. Estimation of signal detection theory parameters from rating-method data: A comparison of the method of scoring and direct search. *Bulletin of the Psychonomic Society*, 1, 207-208, 1973.
- Green, D., & Swets, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.
- Grey, D. R., & Morgan, B. J. T. Some aspects of ROC curve fitting: Normal and logistic models. *Journal of Mathematical Psychology*, 9, 128-139, 1972.
- Kendall, M. G., & Stuart, A. *The advanced theory of statistics* (4th ed., Vol. II). London: Griffin, 1967.
- Krantz, D. H. T. threshold theory of signal detection. *Psychology Review*, 76, 308-324, 1969.
- Luce, R. D. A threshold theory for simple detection experiments. *Psychology Review*, 70, 61-79, 1933.
- Ogilvie, J. C., & Creelman, C. D. Maximum likelihood estimation of receiver operating characteristic curve parameters. *Journal of Mathematical Psychology*, 5, 377-391, 1968.
- Rao, C. R. *Linear statistical inference and its applications* (2nd ed.). New York: Wiley, 1973.
- Restle, F. *Mathematical psychology an introduction*. London: Penguin, 1971.

ORIGINAL PAGE IS
OF POOR QUALITY

APPENDIX F

PRECEDING PAGE BLANK NOT FILMED

COMPUTER TECHNOLOGY

A random search algorithm for laboratory computers

RENWICK E. CURRY

Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

The small laboratory computer is ideal for experimental control and data acquisition. Postexperimental data processing is many times performed on large computers because of the availability of sophisticated programs, but costs and data compatibility are negative factors. Parameter optimization, which subsumes curve fitting, model fitting, parameter estimation, least squares, etc., can be accomplished on the small computer and offers ease of programming, data compatibility, and low cost, as attractive features. A previously proposed random search algorithm ("random creep") was found to be very slow in convergence. We present a new method (the "random leap" algorithm) which starts in a global search mode and automatically adjusts step size to speed convergence. A FORTRAN executive program for the random leap algorithm is presented which calls a user supplied function subroutine. An example of a function subroutine is given which calculates maximum likelihood estimates of receiver operating characteristic parameters from binary response data. Other applications in parameter estimation, generalized least squares, and matrix inversion are discussed.

For many investigators involved in behavioral research, the small laboratory computer is viewed primarily as an experiment monitor, controller, and data acquisition device. The small computer is ideal for this purpose because of its low initial cost and because it is dedicated to the laboratory in which it resides.

The decision is not as clear when considering whether or not to perform postexperimental data processing on the small computer. In favor of the small computer are data compatibility (no tape-to-tape or tape-to-card conversions), low operating cost, and availability. Furthermore, there are many programs designed for these applications (e.g., the DECUS library).

The advantages of processing the experimental data on a large computer are the increases in flexibility, scope, and sophistication of processing techniques attendant with the increase in memory and computational speed. There are many well-documented programs in existence which take advantage of these attributes (e.g., the UCLA BMD series).

A strong case can be made for the small computer in one area of "sophisticated" data processing, and it has received only a modest amount of attention. Parameter optimization, which subsumes curve fitting, model fitting, parameter estimation, least squares, polynomial root finding, etc., is a task which can be performed on the large computer as well as the small. Many algorithms have been proposed (e.g., Fletcher & Powell, 1963) and

several of these are available in coded form (e.g., IBM Scientific Subroutine Package). These algorithms typically require gradient calculations, some require matrix inversions, and all would quickly overwhelm the capabilities of a small computer. As with most algorithms, they only attempt to find one local extremum.

Other algorithms utilize a direct search process and are more readily implemented on a small computer. They are conceptually the most simple: the objective function or cost function $f(\cdot)$ is evaluated at a point in parameter space \hat{x} ; a trial point $\bar{x} = \hat{x} + \Delta x$ is selected, and if $f(\bar{x})$ is an improvement over $f(\hat{x})$, \hat{x} is replaced by \bar{x} . Thus, the best solution is always retained and relinquished only when a better one is found.

Algorithms of the direct search type are differentiated by the manner in which the trial value (or Δx) is chosen. Examples of direct search algorithms are Chandler's (Note 1) STEPIT and Hooke and Jeeves' procedure (1961); these and other methods are compared in Dorfman, Beavers, & Saslow (1973). The performance of these algorithms, as measured by the conventional yardsticks of computing time or the number of times the cost function is evaluated, is usually (but not always) inferior to the more sophisticated approaches. For the small computer, however, time is *not* money, and when the further advantages of less programming effort (e.g., no gradients to be computed) and data compatibility are considered, the direct search algorithms on a small computer become very attractive.

One final feature, and some would say that it is the

This research was sponsored by NASA Grant NGR 22-009-733, Man-Machine Integration Branch, NASA-AMES Research Center.

most important, is that these techniques can more easily search for global, not just local extrema, since the logic for global search is readily incorporated in the direct search logic.

In this paper, we present a direct search program for function minimization which is intended especially for use on small computers. Because of this goal, the primary consideration is small program size rather than speed of computation, number of function evaluations, or computation costs. The program's genesis is the "random creep" algorithm originally proposed by Faureau and Franks (1958) and discussed most recently by Bekey and Ung (1974). After describing the algorithm, we propose a new algorithm (the "random leap") which offers substantially better convergence speed. A FORTRAN program to carry out the random leap procedure is given, and an example of maximum likelihood estimation of receiver operating characteristic parameters is presented. Applications to generalized least squares and matrix inversions are also discussed.

RANDOM SEARCH ALGORITHMS FOR GLOBAL EXTREMA

The Random Creep Method

The random creep algorithm (Bekey & Ung, 1974; Faureau & Franks, 1958) derives its name from the manner in which the trial values of the parameter vector increment Δx are derived: (1) At each stage of the iteration, each of the n elements of the currently optimal parameter vector \hat{x} is perturbed by a zero-mean Gaussian random number of specified standard deviation:

$$\Delta x_j^i = \epsilon_j^i \quad j = 1, \dots, n$$

$$\epsilon_j = N(0, \sigma_j^2)$$

where Δx_j^i is the perturbation to the j th element of x on the i th trial, and the ϵ_j are independent zero-mean Gaussian variables. (2) If the trial vector $\tilde{x} = \hat{x} + \Delta x$ results in an improvement, \hat{x} is replaced by \tilde{x} . If not, a new Δx is chosen according to the above equation. (3) If the iteration fails to improve after a specified number of consecutive trials, all standard deviations are increased by the same factor

$$\sigma_j = \alpha \sigma_j \quad j = 1, \dots, n$$

where $\alpha > 1$. This action is based on the assumption that a local minimum has been reached. (4) The process terminates after either (a) the total number of iterations reaches a predetermined value, or (b) the standard deviations have been increased a specified number of times.

The major advantage of this approach in small computer applications is the relatively small amount of storage space required for the search algorithm. There is

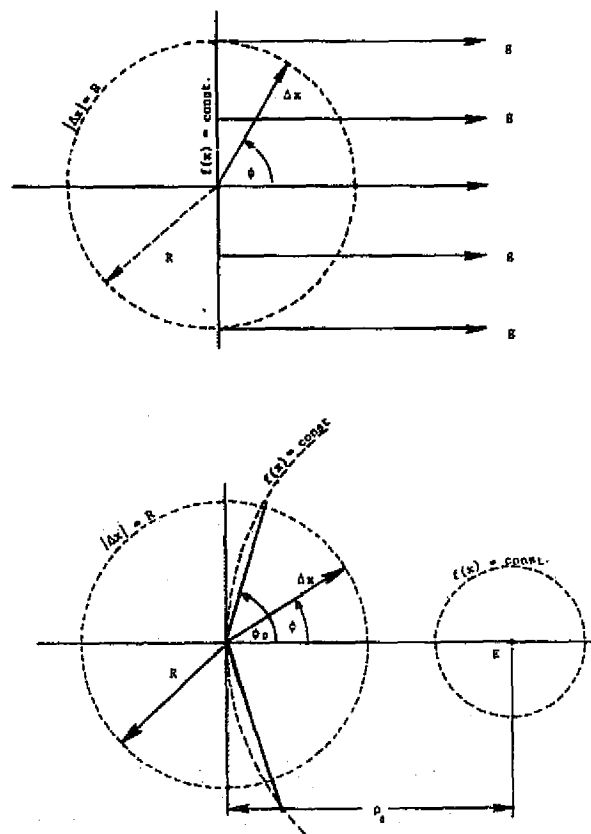


Figure 1. Geometry of possible improvement with random steps (Δx) of fixed size (R) in two-parameter space. (a) Far from the extremum: the gradient vectors (g) are parallel, and an improvement is possible only if ϕ (the angle between Δx and g) is $-\pi/2 < \phi < \pi/2$, which occurs with the probability .5. (b) Close to a spherical extremum (E) at distance ρ_0 : contours of constant criterion $f(x)$ are concentric circles centered at E . Improvement is possible only if $-\phi_0 < \phi < \phi_0 < \pi/2$ with probability $< .5$. When the step size is greater than twice the distance to the extremum ($R > 2\rho_0$), no improvement is possible.

no guarantee that this algorithm will converge to the global extremum (only an exhaustive search will do that), but experience has shown this to be an effective method of finding local and global extrema (Bekey & Ung, 1974).

The Random Leap Method

Improving convergence rate. The main disadvantage of the random creep method, as it is described in the previous section, is the relatively slow rate of convergence because only the minimum step size is used. We experimented with modifications to rectify this problem; for example, if Δx had been an improvement on the previous trial, it was then used in succeeding trials until no further improvement was obtained. These modifications gave a minimal amount of improvement.

It was about this time that the analytical work of Rastrigan (1963) came to our attention; he analyzed a random search procedure where the direction of the

random search vector is uniformly distributed over the hypersphere, but the step size is fixed. The areas of interest are shown in Figure 1 for a two-parameter case: Δx is the random increment in the parameter vector and has a fixed magnitude or step size of R ; ϕ is the angle between the gradient vector g and the random increment Δx . In Figure 1a, the currently optimal value of x is far from the extremum resulting in (nearly) parallel gradients throughout the trial region. The probability of an improvement is $1/2$ under these conditions.

Figure 1b shows the situation when the search procedure nears an extremum E which is assumed to be of a spherical nature—the contours of constant cost function are circles in this two-parameter case, but are hyperspheres in general. In this diagram, the step size is again denoted by R ; ϕ is the angle of Δx from the gradient, ϕ_0 is the maximum angle for which an improvement can be made; and ρ_0 is the distance of the current point from the extremum. It can be seen that the probability of improvement is less than $1/2$ under this condition, since there is a smaller region in which the random search vector will yield an improvement. In the extreme case, when the step size is twice the distance to the extremum, then all points of the function which would yield an improvement lie within the circle of the search vector and no improvement is possible. These factors are shown quantitatively in Figure 2, which

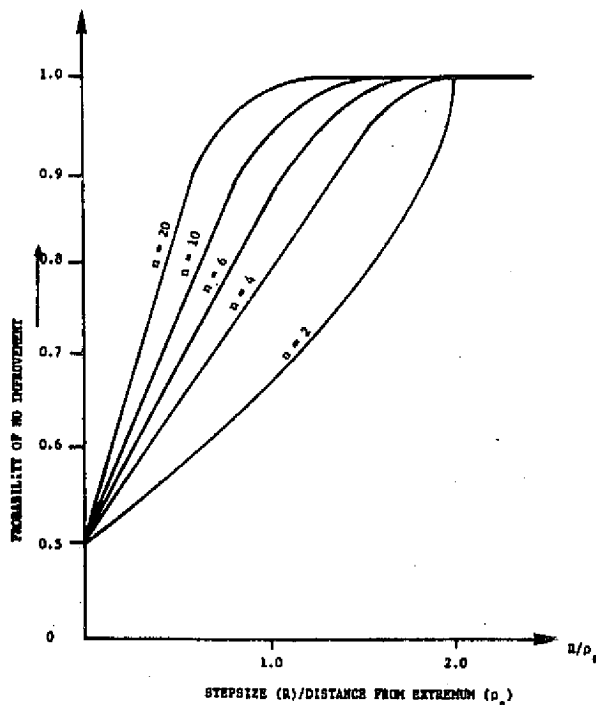


Figure 2. Probability of no improvement vs. step size (measured in distance to extremum) for spherical extremum in n -dimensional space.

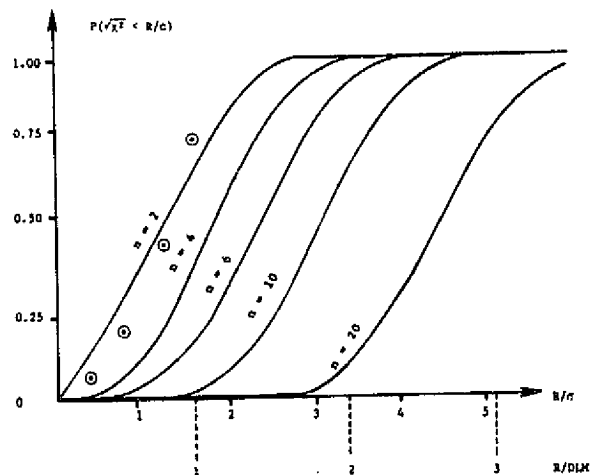


Figure 3. Approximate magnitude R of the random search increment Δx . Upper abscissa scale for R normalized by the standard deviation (Gaussian distribution); lower abscissa scale for R normalized by the half-limit of a uniform distribution ($DLM = \sqrt{3}\sigma$). Circles show exact values for uniform distribution ($n = 2$).

demonstrates the probability of no improvement for fixed step size as a function of the step size, measured in distance from a spherical extremum, for various dimensions of the parameter vector. These curves were generated by numerically integrating the expressions given in Rastrigan (1963).

Figure 2 contains the information that is the basis for the random leap method. In essence, the step size is adjusted so that one always obtains a modest probability for improvement. When the step size is large, it is difficult to get an improvement, so the step size is halved. As the search procedure then comes closer to the extremum, it again becomes difficult to obtain an improvement and the step size is again halved.

When dealing, as we are, with a completely random increment in each coordinate direction, the step size is not fixed. In Figure 3, we show the cumulative distribution curves for the $\sqrt{\chi^2}$ distribution which is the magnitude of the random increment Δx when the increments are independent Gaussian variables with the same standard deviation. The algorithm we suggest uses uniformly distributed variables, so we have also shown, for later use, the abscissa in the scale of step size normalized with respect to the limits of the uniform distribution. The difference between the uniform and Gaussian distributions (in terms of the magnitude of the vector) becomes small for a large number of parameters; the circles in Figure 3 indicate the exact solutions for uniformly distributed variables for $n = 2$, and even for this low number of parameters, the deviations are of no practical significance.

Program outline. The overview of the operation begins with a call to the user-supplied function routine to

```

C*****RANDOM LEAP PARAMETER SEARCH ROUTINE
DIMENSION X(5),TMPX(5),DLMO(5),TMLM(5)
C*****SET J=2**M(P/2)-3,RNDMG=2**M(P-1)-1 FOR P BIT MACHINE
JA = 61
RNDMG = 2047
IRND = 1237
C*****INITIALIZATION CALL TO GET PARAMETERS AND FIRST VALUE
1 LFRST = 1
TCST=CSTFN(TMPX,NX,LFRST,MXITR,MXFGS,MXEXP,MXFLS,MXCTR,LPRT,DLMO)
LFRST = 2
NFAIL = 0
ITER = 0
C*****START LOCAL SEARCH HERE
30 MODE = 1
C*****START GLOBAL SEARCH HERE (MODE SET TO 2 BELOW)
35 NEXP = 0
NCTR = 0
C*****SET TEMPORARY DISTR. LIMITS AT INITIAL VALUE
DO 40 I=1,NX
40 TMLM(I) = DLMO(I)
GO TO (50,80),MODE
C*****SAVE SUCCESSFUL TRY AND PRINT IF LPRNT=2
50 CST = TCST
DO 60 I=1,NX
60 X(I) = TMPX(I)
GO TO (75,70),LPRT
70 WRITE (1,1040) ITER,MODE,NFAIL,NEXP,NCTR,CST,X(1),I=1,NX
1040 FORMAT (516,E15.5,E10.4/(10X,6F10.4))
75 NFAIL = 0
C*****FIND AND TEST NEW TRIAL X AFTER CHECKING FOR MAX ITERATIONS
80 ITER = ITER + 1
IF (ITER - MXITR) 90,90,230
90 DO 99 I=1,NX
IRND = 1A=IRND
IF (IRND) 99,250,99
99 TMPX(I) = X(I) + FLOAT(IRND)/RNDMG*TMLM(I)
TCST=CSTFN(TMPX,NX,LFRST,MXITR,MXFGS,MXEXP,MXFLS,MXCTR,LPRT,DLMO)
IF (TCST - CST) 100,110,110
C*****SAVE IMPROVED SOLUTION BUT FIRST SWITCH TO LOCAL IF IN GLOBAL
100 GO TO (50,30),MODE
C*****OTHERWISE COUNT SUCCESSIVE FAILURES, GO TO STEPSIZE LOGIC
110 NFAIL = NFAIL + 1
GO TO (120,170),MODE
C*****STEPSIZE LOGIC FOR LOCAL SEARCH
C*****TEST FOR MAX FAILURES, LOCAL SEARCH
120 IF (NFAIL - MXFLS) 80,80,130
130 NCTR = NCTR + 1
IF (NCTR - MXCTR) 140,140,160
C*****HALVE DISTRIBUTION LIMITS AND TRY NEW X
140 NFAIL = 0
DO 150 I=1,NX
150 TMLM(I) = TMLM(I)*.5
GO TO 80
C*****END LOCAL SEARCH IF TOO MANY CONTRACTIONS (HALVINGS)
160 WRITE (1,1100)
1100 FORMAT ('END LOCAL SEARCH')
WRITE (1,1040) ITER,MODE,NFAIL,NEXP,NCTR,CST,X(1),I=1,NX
MODE = 2
GO TO 35
C*****STEPSIZE LOGIC FOR GLOBAL SEARCH
C*****TEST FOR MAX FAILURES, GLOBAL SEARCH
170 IF (NFAIL - MXFGS) 80,80,180
180 NEXP = NEXP + 1
IF (NEXP - MXEXP) 190,190,230
C*****EXPAND DISTRIBUTION LIMITS
190 NFAIL = 0
DO 200 I=1,NX
200 TMLM(I) = 1.3*TMLM(I)
GO TO 80
C*****EXIT SEARCH LOGIC HERE AND START AGAIN
230 WRITE (1,1060)
1060 FORMAT ('STOP')
WRITE (1,1040) ITER,MODE,NFAIL,NEXP,NCTR,CST,X(1),I=1,NX
GO TO 1
END

```

Figure 4

supply the constants for the program operation and the initial guess at the parameter vector. From there, the operation proceeds as follows: (a) Uniformly distributed independent increments of each coordinate are chosen to calculate the new trial value of x .¹ Initially, the distribution limits of these random increments are chosen to be of moderate size relative to the entire search region, i.e., if the search region is $(-1,1)$ in each coordinate, then the limits of the uniform increments are on the order of $(-1,1)$. This allows a preliminary global search in the beginning phases. (b) When the trial values of x have failed to yield an improvement a specified number of times, it is assumed that the step size is too large relative to the distance to the extremum, and the distribution limits are halved. The search continues with these limits until the procedure fails to yield an improvement in the (same) consecutive number of trials. (c) After the distribution limits have been

halved a specified number of times, it is assumed that the procedure has converged to a local minimum. The results are printed out, and the distribution limits are reset to their original, moderately sized, values. The global search is then initiated in a manner similar to the random creep method: if no improvement is reached after a specified number of times, the distribution limits are increased, and the search continued. Note, however, that the global search is initiated with the moderately sized distribution limits and not the minimum size, as is done in the random creep method. (d) The procedure is terminated whenever the total number of iterations exceeds a specified value or when the global search has not yielded an improvement after a specified number of expansions of the distribution limits.

The advantage of the random leap algorithm lies in its ability to perform a preliminary global search and gradually reduce the step size as the extremum is neared. There are cases, obviously, when the random creep method would be better, such as when the initial value of x is very close to the extremum.

Program description. A FORTRAN listing of the random leap algorithm is given in Figure 4, which executes the general procedure described above. To do this, it requires a user-supplied function subroutine CSTFN which calculates the value of the cost function; we have made provision for the user to enter the constants and parameters for the search routine for maximum flexibility. During the first call to the function routine CSTFN, the following transfer of variables obtains:

Variables Passed to CSTFN

LFRST has been set to the value of 1 to indicate that this is the first (initialization) call

Variables Returned from CSTFN

TMPX temporary value of X , in this case the initial value

NX the number of dimensions in the parameter vector x

MXITR maximum number of iterations

MXFGS maximum number of consecutive failures in the global search mode

MXEXP maximum number of expansions (increases of the distribution limits) in the global search mode

MXFLS maximum number of consecutive failures in the local search mode

MXCTR maximum number of contractions (halvings) of the distribution limits in the local search mode

LPRT returned as 2 to print (a) each time an improvement is made, (b) at the end of the local search, and (c) at the termination of the search; returned as 1 to print only at the end of the local search and at the end of the entire search

TCST the numerical value of the cost function
 DLMØ the array containing the initial limits of the uniformly distributed random search increments. For example, $\Delta X(1)$ is uniformly distributed between $(-DLMØ(1), DLMØ(1))$.

For subsequent calls to the function routine during the normal course of operation, the following transfer of variables applies:

Variables Passed to CSTFN

TMPX the temporary or trial value of X
 LFRST has been set to 2 to indicate that initialization is not required

Variables Returned from CSTFN

TCST the numerical value of the cost function

CHOOSING THE SEARCH PARAMETERS

In this section, we give some guidelines on how one determines the parameters of the search. Assume that the parameters have been scaled such that the solution is likely to lie within the unit hypercube $(-1,1)$ in each coordinate. Let us pick a convergence criterion that (say) each value should be determined within at least .01 of the true value which minimizes the cost function. To find the minimum step size required, we arbitrarily fix the probability of no improvement which we would like to detect at (say) .7. Looking at Figure 2, we can then determine the step size relative to the distance from the extremum for various dimensions of the parameter vector. If we have four parameters, then a probability of .7 occurs with a step size that is roughly .6 times the distance from the extremum. Since we want the distance from the extremum to be no more than .01, the magnitude of the minimum step size must be .006. To find the minimum distribution limits corresponding to the minimum step size, we refer to Figure 3, where we see that for $n = 4$, the step size will be less than 1.8 times the distribution limit more than 95% of the time. Thus, the minimum limits of the uniform distribution should be $\pm .006/1.8 = \pm .0033$. The initial value of the distribution limits should be on the order of ± 1 , and because the limits are halved at each stage, we seek an integer M such that $1/2^M \approx .0033$. Note that $2^8(.0033) = .768$, so we set $DLMØ = .8$ and $MXCTR = 8$.

The adaptive nature of the random leap program results from testing the hypothesis that the probability of failure is .7 or less. To reject this hypothesis at the .05 and .01 levels of significance, we would need 9 and 13 successive failures, respectively. (six and nine successive failures are required for the same level of confidence when detecting probability of failure .6 or less, showing the desirability of working with probabilities of no improvement closer to .5 than 1.0.) Thus, $MXFLS$, Maximum Consecutive Failures, Local

Search, should be in the range of 9 to 13, perhaps greater to account for contingencies.

The parameters for the global process ($MXFGS$ and $MXEXP$) can really only be determined by experimentation in each particular situation. We found, in the cost function described below (which has two maxima) that 300 trials with a five-dimensional vector were more than adequate to find the global maximum when situated at the local maximum. The number of such trials must be increased substantially when going to a larger parameter vector to obtain similar densities of trial values within the parameter space.

APPLICATIONS

Maximum Likelihood Estimation

In this section, we describe an application of the random leap algorithm and the results of a Monte Carlo simulation. The first example uses a cost function of the form

$$f(x) = .59\exp(-Q_1(x)) + .41\exp(-Q_2(x)) \quad (1)$$

$$Q_1 = S(n) \sum_{i=1}^n (.5 - x_i)^2 / (1.26)^{i-1} \quad (2a)$$

$$Q_2 = S(n) \sum_{i=1}^n (.5 + x_i)^2 / (1.26)^{i-1} \quad (2b)$$

where $S(n)$ is a scale factor depending only on the number of dimensions in the parameter vector and keeps the argument of the exponent function within reasonable bounds. This function was chosen for a detailed examination because of its two nearly equal modes. It corresponds to a maximum likelihood estimation problem in which the observations $(.5, .5, \dots)$ come from either a distribution with mean x (prior probability .59) or $-x$ (prior probability .41) and unequal variances.

Monte Carlo trials. One hundred Monte Carlo trials of the random creep and random leap operations were run using the cost function in Equation 1 on a PDP-12 computer with software multiply and divide routines. Approximate time for each iteration was about 1.5 sec. The initial conditions for these trials were uniformly distributed in the unit hypercube, and parameter vectors of dimensions 2 and 5 were used. Convergence was considered complete when the search routine first came within a sphere corresponding to an rms deviation in each component of .05.

With the two-dimensional parameter vector, the mean number of iterations-to-convergence for the random leap algorithm was 35, the median 36, and the maximum number of iterations required was 67. In all these cases, the fact that the search was started at the global level rather than the local level resulted in a convergence to the global maximum first. On the other hand, the random creep algorithm converged to the local

maximum first in many trials. It also failed to find the global maximum within the allowable maximum of 500 iterations on many of its trials, whereas the random leap algorithm never failed to converge to the global maximum.

In testing the random leap algorithm with the five-dimensional parameter vector, the mean number of iterations-to-convergence in 100 Monte Carlo trials was 194, with a maximum number of iterations of 365 and a minimum of 58. The distribution of the number of iterations had two modes: one, at 100 iterations, corresponded to those searches that went directly to the global maximum; the other mode, at 240 iterations, represented those searches that went to the local maximum first and then to the global maximum. Sixty of the 100 Monte Carlo trials converged to the global maximum first; this is a statistically significant difference from the expected value of 59 that one would expect with the random creep algorithm, and is due to the fact that the random leap algorithm starts off in the global search mode. The random creep algorithm was not run in this more difficult task because of its relatively poor performance on the easier two-parameter case.

Receiver Operating Characteristics

In the theory of signal detection (TSD) approach to psychophysics, the subject's response is divided into sensory and response bias components. One may administer a yes-no response procedure in several sessions, during each one of which a subject adopts a different strategy or criterion level (β), whereas the sensitivity of the signal (d') remains constant throughout (Green & Swets, 1966). Assuming that we have binary response data ("YES" or "NO") for M criterion levels, the TSD model uses the following expression for the probability of a "YES" response given noise (n) and signal(s) respectively

$$P(\text{"YES"}|\text{noise}) = 1 - \Phi(\beta_i) = \Phi(-\beta_i) \quad (3)$$

$$i = 1, 2, \dots, M$$

$$P(\text{"YES"}|\text{signal}) = 1 - \Phi[(\beta_i - d')/\sigma_s] = \Phi[(d' - \beta_i)/\sigma_s] \quad (4)$$

where β_i is the criterion level adopted in the i th session, d' is the sensitivity measured in noise standard deviation units, σ_s is the standard deviation of the signal distribution, measured in noise units and Φ is the Gaussian distribution function.

The maximum likelihood estimates for the parameters β_i , d' , and σ_s are obtained by maximizing the likelihood function which is proportional to

$$\ell(x) \propto f(x) = \sum_{i=1}^m r_{iN} \ln[\Phi(\beta_i)] + r_{iY} \ln[\Phi(d' - \beta_i)]$$

$$+ r_{iN} \ln[\Phi(\beta_i - d')/\sigma_s] + r_{iY} \ln[\Phi(d' - \beta_i)/\sigma_s] \quad (5)$$

where r_{ijk} is the number of j responses at criterion i to stimulus k , $i = 1, 2, \dots, M$; $j = \text{YES, NO}$; $k = s$ (signal), n (noise). The parameter vector for this case is

$$x = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \\ d' \\ \sigma_s \end{bmatrix} \quad (6)$$

Note that we can easily obtain the parameter estimates under the constraints of equal-variance distributions by setting σ_s in the above expression to unity.

A function subroutine written to generate the parameter estimates is shown in Figure 5. It prompts the user for the number of different criterion levels, and a parameter indicating whether it is to be an equal-variance case or whether $B = 1/\sigma_s$ is a parameter to be estimated.

```

FUNCTION CSTFNCK,NX,LFRST,MXTR,MXF,MXEXP,MXFLS,MXCTR,LPRTR,DLMO)
C*****FUNCTION SUBROUTINE FOR MAXIMUM LIKELIHOOD ESTIMATION OF ROC PARAM
C*****WRITTEN FOR TELETYPE INTERACTION
DIMENSION X(1),RC(2,6),DLMO(1)
GO TO (1,30),LFRST
C*****THIS PORTION FOR THE INITIALIZATION CALL TO INPUT DATA
1 WRITE (1,1000)
1000 FORMAT (' RE:15 NX,MXTR,MXFGS,MXEXP,MXFLS,MXCTR,LPRTR')
READ (1,1010) NX,MXTR,MXFGS,MXEXP,MXFLS,MXCTR,LPRTR
1010 FORMAT ('34')
WRITE (1,1005)
1005 FORMAT (' NUMBER OF CRITERIA/1 FOR EQUAL VAR,2 OTHERWISE')
READ (1,1010) MCRT,LB
IF (X(LB-1)*(LB-2)) 1,5,1
5 WRITE (1,1020)
1020 FORMAT (' READ #NO,YES RESPONSES;FIRST NOISE, THEN S + N')
DO 10 ICRIT = 1,MCRT
DO 10 ISIG = 1,2
1 = ISIG + 2*(ICRIT - 1)
10 READ (1,1030) RC(1,I),RC(2,I)
1030 FORMAT ('F10.2')
WRITE (1,1040)
1040 FORMAT (' READ X(C),DLMO(I)')
IMX = MCRT + LB
DO 20 I = 1,IMX
20 READ (1,1030) X(I),DLMO(I)
C*****NORMAL CALCULATIONS HERE
C*****STRIP D' FROM X AND LIMIT IF NECESSARY
30 IX = MCRT + 1
IF (X(IX) - .01) 34,38,38
34 X(IX) = .01
38 D = X(IX)
C*****GET STD. DEV. RATIO
GO TO (40,50),LB
40 B = 1.
GO TO 60
50 IX = MCRT + 2
C*****LIMIT ALLOWABLE B .GE. 0.01
IF (X(IX) - .01) 54,58,58
54 X(IX) = .01
58 B = X(IX)
C*****CALCULATE LOG LIKELIHOOD FUNCTION (LESS CONSTANTS)
60 XLLF = 0.
DO 100 ICRIT = 1,MCRT
DO 100 ISIG = 1,2
GO TO (70,80),ISIG
70 Z = X(ICRIT)
GO TO 90
80 Z = B*(X(ICRIT) - D)
C*****THEORETICAL PROBC('NO') FROM GAUSSIAN CUMULATIVE DISTRIBUTION
90 P = GCDP(Z)
IP = ISIG + 2*(ICRIT - 1)
100 XLLF = XLLF + RC(1,IP)*ALOG(P) + RC(2,IP)*ALOG(1. - P)
C*****MINIMIZE THE NEGATIVE LOG LIKELIHOOD FUNCTION
CSTFN = -XLLF
RETURN
END

FUNCTION GCDP(Z)
AZ = ABS(Z)
T = 1./((1. + .2316419*AZ)
D = .3989423*EXP(-Z*Z*.5)
P = 1. - D*F((C(1.330274*Z - 1.821256)*T + 1.781478)*T
O = .3363636)*T + .3193815)
IF (Z) 1,2,2
1 P = 1. - P
2 GCDP = P
RETURN
END

```

Figure 5

For a test case, we assumed a value of x equal to $(\beta_1 = .5, \beta_2 = .75, \beta_3 = 1.25, d' = 1.00, \sigma_s = 1.25)$. We used as responses the values proportional to the theoretical probabilities, i.e., we expect a "perfect" fit. Starting from the initial condition $(-.5, 0, .5, .5, .5)$, the random leap algorithm converged to within .01 of the true value of each component of the parameter vector within 100 iterations. To determine how well these data from an unequal variance distribution are fit by an equal-variance model, we used the equal-variance option of the program and found the equal-variance fit to the unequal-variance data; the final estimates were $(\beta_1 = .531, \beta_2 = .756, \beta_3 = 1.1, d' = .96)$, and the random leap algorithm, starting from the same initial conditions as before, again converged to within .01 of the final values after 100 iterations. Each of these solutions took about 3 min on the PDP-12 computer described above. Since a printout was made at each improvement, though, the computations were limited at times by the print operation.

Parameter Estimates from Grouped Data

The maximum likelihood procedure is one method of obtaining estimates of distribution parameters for grouped data, for example, obtaining the mean and standard deviation of a distribution from data such as is gathered in poststimulus histogram form. In these cases, the log likelihood function is proportional to

$$\ell(x) \propto \sum_{i=1}^N r_i \ln P_i(x) \quad (7)$$

where r_i is the number of responses in the i th group ($i = 1, 2, \dots, N$), and P_i is the theoretical probability of the sample falling in the i th group. Under the assumption that the underlying distribution is Gaussian, P_i is given by

$$P_i = \Phi[(c_{i+1} - \mu)/\sigma] - \Phi[(c_i - \mu)/\sigma] \quad (8)$$

$$c_1 = -\infty, c_{N+1} = +\infty$$

and the parameter vector is

$$x = (\mu, \sigma) \quad (9)$$

Simple changes can be made to the function routine shown in Figure 5 to perform these calculations. The boundaries of the groups c_i must, of course, be read in during the initialization phase, but it is a simple matter to program Equation 7 and solve for the mean, μ , and standard deviations, σ .

An alternative to using the maximum likelihood estimation criterion is the minimum χ^2 criterion (Rao, 1973). This has many of the advantages of the maximum likelihood procedure, and serendipitously, one is

calculating the χ^2 value which can be used in a goodness-of-fit test when the final estimates have been obtained.

Generalized Least Squares

One of the most common and most powerful forms of regression analysis is generalized least squares, which can be written in the form

$$f(x) = (\tilde{y} - h(x))^T W (\tilde{y} - h(x)) \quad (10)$$

where \tilde{y} is a k -dimensional vector of observations (data), h is a vector function of the parameter x , and W is a positive semidefinite (symmetric) matrix of weighting coefficients. The objective is to choose the parameter vector x to minimize this cost function. In the linear form, Equation 10 becomes

$$f(x) = (\tilde{y} - Hx)^T W (\tilde{y} - Hx) \quad (11)$$

where H is a $k \times n$ matrix. This has a unique solution under suitable conditions on H and W (Rao, 1973), usually satisfied in practice. One may use the random leap algorithm to find the solution to Equation 11 rather than go through the matrix inversions and manipulations required to solve for the vector x . We also note that if H is square and nonsingular, and if we solve Equation 11 with \tilde{y} equal to zero except for a 1 in row i , then solution x is the i th column of the matrix H^{-1} . Thus, n separate parameter searches will completely invert the $n \times n$ matrix H .

CONCLUSIONS

The disadvantage of a small laboratory computer for postexperimental data processing is its small memory and inability to accommodate large sophisticated data processing programs, while the advantages of using the small computer are data compatibility, ease of program development, ability to run for long periods of time, low cost, and accessibility. Parameter optimization is one area which has received relatively little attention in small computer applications. Direct search methods are the easiest to program, and they require little core storage and no analytical gradient calculations.

Random search algorithms are one member of this class, and Rastrigan (1963) has shown that the average rate of convergence of the random creep algorithm may actually be superior to the gradient method in which the gradients are calculated numerically at each step. The slow rate of convergence experienced in the random creep algorithm is due to small step size and was overcome by the random leap algorithm proposed here. This algorithm operates by starting off in the global search mode and automatically reducing the step size as the search procedure approaches the extremum of the function. After reaching a minimum (or maximum), it

branches out and starts a global search by gradually expanding the size of the random increments from their original, moderately sized values.

The random leap algorithm was compared to the random creep method on a bimodal likelihood function and showed superior convergence characteristics. We then presented a function subroutine to be used with the random leap program to calculate the maximum likelihood estimates of receiver operating characteristic parameters in YES-NO tasks; this was followed by a description of maximum likelihood estimation of distribution parameters from grouped data (histograms). Generalized least squares regression was also considered, and it was shown how one could perform matrix inversion using successive applications of direct search methods.

REFERENCE NOTE

1. CHANDLER, J. P. *Subroutine STEPIT: An algorithm that finds the values of the parameters which minimize a given continuous function*. A copyrighted program. J. P. Chandler, Copyright, 1965.

REFERENCES

- BEKEY, G. A., & UNG, M. T. A comparative evaluation of two global search algorithms. *IEEE Transactions on Systems, Man and Cybernetics*, 1974, SMC-4, 112-116.
- DORFMAN, D. D., BEAVERS, L. L., & SASLOW, C. Estimation of signal detection theory parameters from rating method data: A comparison of the method of scoring and direct search. *Bulletin of the Psychonomic Society*, 1973, 1, 207-208.
- FAUREAU, R. R., & FRANKS, R. G. Statistical optimization. In: *Proceedings of the Second International Computer Conference*, 1958, 437-443.
- FLETCHER, R., & POWELL, M. J. D. A rapidly convergent descent method for minimization. *The Computer Journal*, 1963, 6, 163.
- GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.
- HOOKE, R., & JEEVES, T. A. Direct search solution of numerical and statistical problems. *Journal of the Association for Computer Machinery*, 1961, 8, 212-229.
- RAO, C. R. *Linear statistical inference and its applications* (2nd ed.). New York: Wiley, 1973.
- RASTRIGAN, L. A. The convergence of random search method in the presence of noise. *Automation and Remote Control*, 1963, 24, 1337-1342.

NOTE

1. This is easily done by intentionally overflowing the FORTRAN integer multiplication, and on a 12-bit machine, provides 1,024 uniformly distributed random numbers (-2047, 2047) before repeating the sequence.

(Received for publication November 27, 1974;
revision received April 1, 1975.)